It's Complicated:

Rethinking Education Reform & Accountability at the Intersection of Federal, State and Local Policy

Christina M. Breitbeil

TC 660 H

Plan II Honors Program
The University of Texas at Austin

Dr. Anthony Petrosino
Curriculum & Instruction
Supervising Professor

Dr. Pedro Reyes
Educational Administration
Second Reader

Acknowledgements

Thank you to my wonderful first and second readers, Dr. Petrosino and Dr. Reyes. I, quite literally, could not have completed this thesis without your help.

Thank you to my amazing mother, for staying on the phone with me as I walked home from coffee shops at 4 am (after the many late nights spent completing this work).

Thank you to my friends, for enduring my many education and thesis related rants, and remaining my friends.

Thank you to the students at both LBJ and Reagan Early College High Schools, for being the one of the primary sources of my interest in this topic, and for always loving to learn.

Table of Contents

Abstract

This thesis demonstrates that successful education reform can be achieved only by a concurrently implemented policy that addresses issues in education which focus on quality and equity at the local, state and federal levels of government. This paper shows that current student performance data provides evidence for the failure of previous reforms that focus on a broad-reaching simplistic approach to reform focused on high stakes accountability finance and testing. Building on this evidence and on the existing literature discussing accountability, this thesis proposes five principles for education policy. However, the complex system of decentralized governance in the United States demands the coordination between federal, state and local lawmakers to draw cohesive policy at each level that will combat the nationwide achievement gap while being cautious of over testing.

A methodological approach utilizing literature review, qualitative and quantitative analysis of legislation, and evaluation of student performance data was used to propose multiple policy recommendations (PRs) at the federal, state and local level. The proposed reform focuses recommendations on enabling lower stakes systems of accountability. Further, it demonstrates how the three coordinating pairs of reformative policies provide both the underlying foundation and overarching structure necessary for each individual policy recommendation to achieve success. Finally, the paper discusses synchronization between each level of government and explains how the five policy recommendations constituting this proposal for reform will best facilitate an educational environment providing every student with equal opportunity for achievement.

I. Introduction

i. Introduction & Summary of Findings

Policy makers on both sides of the political spectrum have for years pushed accountability as the gold standard in education policy. Standards-based reform has become the mainstream effort in terms of "fixing" the system since it was introduced in the 1994 reauthorization of ESEA as the Improving America's Schools Act (IASA). Since then, ESEA has been reauthorized in NCLB with federally prescriptive standards and punitive sanctions for accountability and in ESSA with sanctions but no specifically prescribed standards. In terms of standards-based policy, the US has implemented several variations on legislation in attempt to incentivize states and districts to set and meet standards, without significant improvement in student performance and almost no progress on closure of the achievement gap (Guisbond, et. al, 2013, p. 7). In many countries across Europe and Asia, standards-based policy has also become the norm in education, with significantly more success than the United States.[1]

There are several reasons for the success of standards-based reform in European and Asian countries that did not similarly appear in the US or many of our states. First, in many of these countries, homogeneity in economic security is conducive to a larger portion of the country's students having consistent access to resources that aid in educational achievement. For example, Finland, which is oft cited as a model of educational achievement for U.S. systems to imitate, has a population of 5.3 million compared to the U.S.'s 318 million, a more dominant and narrow heritage (Finnish), and speaks three languages compared to the over 300 spoken in the U.S. (Shumer 2014). The second reason, and what will be discussed at length for the remainder of this

[1] In 2009, Shanghai participated in the Program for International Student Assessment (PISA), league tables of comparison administered by the Organization for Economic Cooperation and Development (OECD), and ranked first in all three categories of assessment (Robinson 2015, 7).

section, is that the structure of policy in European and Asian countries with successful standards-based reform is far simpler than the unique structure of United States policy. In contrast to other centralized nations, the interaction of federal, state and local oversight and simultaneous decentralization of policy creates a tug between interstate cohesiveness and independent state constitutions in the evaluation of US performance in education. Sweeping policy measures such as high-stakes accountability, whether implemented at the top by the federal government or throughout as within many state governments, have not produced their intended effects on performance or equity (Bussert-Webb 1999, Marker 2001, Koza 2002, Berliner and Nichols 2005, Moses and Nanna 2007, Mintrop and Sunderman 2009, FairTest 2010, Phelps 2011).

The stock of data building since the early 1990s as standards-based reform – and standardized testing – gained steam across the nation has provided the evidence that confirms validity of the EPPs discussed in Part I. Consequently, lawmakers must construct future policy with these principles in mind. There is much to learn both from the failure of past reforms and from the valuable student performance data that they nonetheless produced despite not meeting goals. Accountability – in the general sense – as a means to safeguard the right to sufficient education for all US students is a necessary, equitable element of the education system. Accordingly, some version of standards under this system of accountability must apply to all students equitably; every student's performance must be incorporated into the data (EPP 3). Low, rather than high, stakes should be attached to these standards, to ensure that punitive stakes do not incentivize unequal treatment of low-performing students in order to preserve higher performance (EPP 4). A productive system of standards can be achieved, to a limited extent, by testing students and scoring standardized tests to measure student performance. However, this cannot be the only indicator of student performance: other, non-standardized methods of assessment must be used in conjunction with standardized tests in order to minimize the

corruptibility of data (EPP 5). Alternative assessments that both facilitate and measure higher-order thinking best serve as this additional indicator, and this is achieved by holistically measuring performance on project-based, analytical tasks (EPP 2). A productive system of standards for the purposes of *funding* must be based primarily on need, rather than merit: accordingly, demographics must be included in the formula for comparing school performance to determine allocation of funds (EPP 1). So how do we satisfy all these principles from the policy perspective?

The complicated and intricate nature of policy in the United States requires that future reform begin to account for the intersecting areas of federal, state and local policy measures of reform at the top are unable to account for the protections of every student as it trickles down through state, local, and individual school jurisdictions. Yet, measures in education policy at the federal level are necessary to ensure progress is made to close the achievement gap across the nation. The current system in place includes: heavy regulation and mandates for high-stakes at the federal level, accountability and standards set at the state level, and additional layers of accountability and bureaucratic methods for improved achievement at the local level. The stakes set at the federal level currently impose narrowly defined guidelines for states to follow regarding accountability, and though states set the standards and create the tests, directives at the federal level ensure the weight given to those tests – impacting policy at the state and local level in profound ways. As a result, the current education policy at the federal level leaves little room for states to pursue productive, new methods of accountability – regulations at the federal level require stakes at the state level, and from the local (and state) perspective, the stakes are simply too high to deviate from the norm. The problem, however, is that this norm – high-stakes accountability based on state standardized test scores – is ineffective, and, increasingly, harmful (Bussert-Webb 1999, Marker 2001, Koza 2002, Berliner and Nichols 2005, Moses and Nanna

2007, Mintrop and Sunderman 2009, FairTest 2010, Phelps 2011).

The reforms did produce something of value: data. The question becomes, then, how should policymakers use analysis of past reforms and student performance to rectify their consequences? The first area demanding action, and on which this paper will primarily focus, is reform at the federal level: The stakes set at the federal level must be lowered, and much of the narrow policy regulations must widen to provide room for new forms of accountability. Because the top-down regulations currently in place are the primary obstacles to progressive education reform at the state and district levels, the policy recommendations made in this paper for implementation at the federal level are largely critical in nature. Policy Recommendation 1 (PR 1) is to reauthorize ESEA with a new set of legislation lowering and reversing many of the stakes set in the current reauthorization under ESSA. Consequently, the argument for this is largely based in a critical analysis of the mandates and stakes set in ESSA. Policy Recommendation 2 (PR 2) is to eliminate Race To The Top (RTTT), and to replace it with a grant that encourages and rewards States for improvement under systems of "intelligent accountability" – a flexible concept embodied by accountability systems that facilitate minimal student exposure to standardized tests and maximum stimulation of higher-order learning. Again, for this reason, the argument for a new federal grant program consists largely of evidence for the failure of the current standards-based RTTT grant.

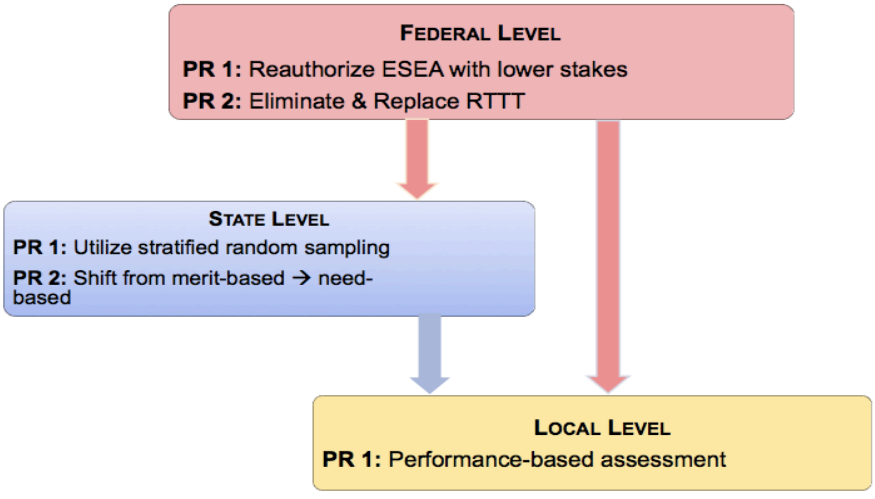
Making these improvements at the federal level subsequently allows implementation of reforms at the state level. Because the recommendations at the state and local levels do not focus on a specified governing (such as a specific state or district), the policy recommendations made in this paper for implementation at the State level are general models for reforms that should be implemented with consideration of the specific needs of constituents. State PR 1 is to utilize a system of stratified random sampling in order to produce data on student performance without

testing every student every year. The argument for this recommendation follows that random sampling would – while producing accurate data across demographics – minimize classroom incidence of "teaching to the test" by lowering stakes tied to student performance on standardized tests and reducing individual student exposure to standardized assessments. State PR 2 is to incorporate demographics based on need – such as percentage of low-income students – in formulas to determine allocation of funds to schools. By doing this, states would move from a merit-based system of funding to a need-based one, which has demonstrable benefits for equity in education.

Finally, the flexibility at the state level to implement these reforms without the stakes underlying immediate improvement will open the door for LEAs to introduce performance-based assessments. These assessments are project-based, analytical and autonomous in nature, which enables the assessments to promote rather than hinder higher-order learning and twenty first century skills.

Figure 1: Policy Recommendations (PRs)

The table below lists the five policy recommendations (PRs) made in this paper, at various levels of policy. Note that each of the levels of policy recommendations are connected by an arrow, which indicates the intersectional nature of the PRs. PRs at the federal level facilitate PRs at the state and local levels, and PRs at the state level facilitate the PR at the local level.



Examination of individual policies reveals that each has broader implications for varying levels of governing authority, at the federal, state and local levels. This has indicated that concurrently implemented policy must address issues in education at every level U.S. education policy: for progressive measures to succeed at the district level, the adequate policy must be in place at the state level, and for policy measures to succeed at the state level, federal policy must be conducive to rather than preventative of the success of those measures. The same statement could be made in the reverse.

The layout of this structure, in part, addresses the means for implementation of this reform. An ideally functioning system would include a district, state and federal government, each of which implements these reforms according to the particularities of individual circumstances at each level. By addressing policy on every level and ensuring that reformative measures each cohesively build on one another, the broad proposal provides both the foundational and overarching elements necessitated by a system conducive to success for *all students.*

ii. Methodology

Research Questions

In relative terms, the focus of this thesis is very broad. This is, in large part, because it attempts to address the policy issues inherent to an area of mammoth scope: the United States education system. Ultimately, this thesis attempts to provide a framework – and subsequent individual policy recommendations – for a successful education system. At the center of the discussion on education and relevant policies in this paper is the subject of accountability. Two points have informed the paper's consideration of this issue. The first is that accountability seems to be increasingly necessary in a system that must constantly progress toward an equitable and

beneficial education system; the second is that past and current systems do not appear to achieve this for all students, and have produced many unintended consequences. The questions this thesis attempts to answer subsequently develop from these points:

1. How can policy hold institutions of education accountable for ensuring equal opportunity and quality of education for all students, without creating a culture of over-testing and rote learning?
2. How can reformative policies achieve this at the intersection of federal, state, and local levels of government?

The first question has been pursued in varying levels of specificity by numerous pieces of research, and has been a particularly popular topic since the origins standards reforms in the 1970s. Complications in answering this question stemmed from the breadth in quantity of the relevant bodies of research, and the resulting task: dissecting the volume of the topic to identify the most valuable elements from a policy perspective.

The second question presented more extensive complications. Investigating the impact of policies from the student perspective revealed the complicated nature of their individual and interacting effects. Broad federal authority, particularly in measures to direct accountability, in a simultaneously decentralized system of education in which power lies at the state and local levels to produce intersecting lines of political intervention. Recognizing this indicated that no one level of government could produce a sweeping policy reform to solve the issues in education.

Structuring Findings

Surmounting complications to answer these questions required strategically structuring both the research for this paper and the findings of the paper itself. During the research process, analyses of information required differentiation between findings with implications for broad principles of policy at every level, and findings with implications for future policy proposals at

specific levels of government. Analyses falling under the former categorization are contained within Part I of this thesis, which identifies five broad principles of education policy (EPPs), each of which may apply to various levels of government. Analyses falling under the latter categorization are contained within Part II of this thesis, which presents and defends individual policy recommendations (PRs) at specific levels of government: this includes two PRs at the federal level, two PRs at the state level, and one PR at the local level. Evaluation of research for Part II of this thesis required an additional element of analysis: Due to the intersectional nature of the recommendations, research of individual policies necessitated synthesis across various levels of governing. For example, policies identified for recommendations at the federal level must facilitate those recommended at the state level, and state policy recommendations must coordinate with policies at the local level.

The immense scale of the task undertaken by this thesis – to analyze, propose, and integrate legislation across all levels of government – demanded an appropriately scaled approach. Accordingly, this thesis has become an evaluation of and recommendations for policy at *ten thousand feet*. For this reason, only policy recommendations at the federal level had a specific focus of scope – the focus being United States policy; no individual states or districts are specified for direct analysis. Consequently, a large portion of policy analysis occurs in the chapter on federal PRs: because the specific governing body is identified in addition to the level, specific legislation required be critical analysis before new recommendations could be proposed.

Variables of Consideration

Throughout research to determine principles of education policy and recommendations for future policy measures, three components were used throughout the research process to frame analysis of policy. The first variable of consideration was equity. Given that equal

opportunity is an essential element of the United States' democratic system of government, assessment of the validity and soundness of policies was framed primarily within each policy's maintenance of equity and progress toward a more equitable system. The next variable of consideration was quality of education. This is the purpose of the education system – to not only provide educational services, but to provide services of quality and value. Evaluating within the frame of this variable required first establishing a foundational understanding of quality of education itself. To subsequently assess this variable, the author's analysis of bodies of research was centered on distinguishing elements of educational practices that best facilitated student development of necessary and beneficial skills. The final component constituting this paper's primary variables of consideration was policy feasibility. Neither equity nor quality of education is of consequence in policy considerations, if the policy in question cannot be feasibly implemented and followed. Thus, past policies were evaluated with consideration of their success in this regard, and proposed policies were similarly evaluated to determine utility, both independent from and intersecting with other policies. Accordingly, the research presented in this thesis will consistently refer to the variables of equity, quality of education, and policy feasibility in consideration and analysis of various findings.

Methods of Research

Literature Review

The most significant method utilized to produce the findings within this thesis was a thorough and in-depth review of existing literature. Diagnosing issues in the current education system in order to identify and assemble the five EPPs relied solely on this method. This included extensive review of works that analyzed the implications of student performance data, and particularly those that tied those implications to specific policies. Reviewing the literature in this manner

allowed the author to make determinations about the adequacy and consequences – both harmful and beneficial – of policies on future student performance. Various sections of this paper pull from certain studies more emphatically than others. For example, the author uses primarily Berliner and Nichols' (2005) body of work on Campbell's Law to defend EPP 5, which presents the necessity of multiple indicators in systems of accountability.

Qualitative Analysis of Legislation

Because this thesis aims to present a framework of successful policies for implementation in future legislation, a valuable method in the course of the research was qualitative analysis of the varying impact of current legislation. The author utilized this method to assess the specific nature of the stakes set at the federal level under current policies. This included an in-depth analysis of the guidelines outlined in the Every Student Succeeds Act (ESSA), with particular attention placed on the retention of stakes at the federal level despite the legislation's replacement of the No Child Left Behind Act (NCLB), which legislated broad federal powers. The other component of the author's usage of this method included synthesizing analysis of policies at varying levels in order to determine the culminating effects of federal, state, and local policies at the intersection of each of the three levels of government authority. This included analyzing the limits placed by federal legislation on state and local policies, and the limits placed by state legislation on local policies.

Quantitative Analysis of Legislation

The author's use of this method took place exclusively within the section of Chapter IV entitled, "Federal Policy Recommendation 2." In employing this method of legislative analysis, the author quantified the weight system used in Race To The Top (RTTT) for federal determination of States eligible for receipt of the grant funding. By identifying the measures on which RTTT

placed the heaviest weight and analyzing the stakes underlying those measures, the author was able to made determinations about the value system at the federal level with regard to this policy.

Analysis of Public Student Performance Data

The final method used in the research for this thesis was an analysis of public student performance data. The author analyzed changes in student performance to make inferences about the simultaneously occurring policies that potentially impacted these performance trends. In particular, the author aligned trends in the achievement gap from the 1960s up until 2013, as indicated by student performance, with the shift from need- to merit-based policies that occurred during that time period to draw conclusions about the effects of those policies. Additionally, the author analyzed the soundness of initiatives such as the New York Performance Standards Consortium by assessing effects the impact of the program on student performance.

iii. Definition of Terms

The terms below are defined by the author in terms of their usage within this body of work:

Education policy – The institutionalized system of total policies governing any action within the education sector. This action may occur at multiple levels of authority: federal, state or local (under Local Education Agencies).

Federal policy – The legislation implemented at the federal level that articulates the specific authorities of the federal government with regard to education, either at the national, state, or local level. Includes oversight of state and local policies.

State policy – The legislation implemented at the state level that articulates the specific authorities of a state government with regard to that state's education, either at the state or local level. Includes oversight of local policies.

Local policy – The legislation implemented at the local level that articulates the specific authorities of Local Education Agencies (LEAs) or districts with regard to education. Includes oversight of school policies.

Accountability – The practice of holding institutions of education responsible for the performance of students, by in some way assessing performance and progress of performance.
Systems of accountability – The methods and processes employed by governing bodies in the practice of accountability.
Standards – The set objectives for student performance and achievement. Typically associated with practices to ensure meeting of or progress toward standards.
Stakes – Any consequences of student performance that significantly impact students, teachers, principals, administrators, schools, districts, LEAs, or states. Most often associated with consequences for students, such as continuation to the next grade level.
High stakes assessment – Any assessment that includes stakes associated to student performance.
Assessment – Any method employed to measure student performance and achievement.
Performance – Formal (as in measured and assessed) demonstration of academic ability.
(**Performance**) **Indicators** – Any measurement used to inform formal decisions (usually at governing levels) in systems of accountability. Measurements may be of student, teacher, school or district performance.
Achievement gap – The significant (documented) difference between the performance of minority students, such as African American or Hispanic students, and the performance of white students, attributed to differences in access to opportunities. Minority students tend to rest at the lower performing end, on average, while white students tend to rest at the higher performing end, on average.
Intelligent accountability – Systems of accountability that rely on alternative methods to measure student performance, that both lower stakes associated with performance and facilitate student development of higher order learning skills.
Performance based assessments – Assessments in which student performance is evaluated based on performance on a specified task, and clear expectations have been outlined for the student preceding the task. Typically project-based in nature and subject focused.
Education Policy Principle (**EPP**) – A broad, fundamental principle (as in rule) of education policy that should be addressed by future policies. Informed by analysis of past and present policies.

Policy Recommendation (**PR**) – A proposal of a specific policy measure for implementation at one level of either the federal, state, or local governments. PR will address one or more EPPs, and coordinate with all other PRs at varying levels of government.

PART I

PRINCIPLES FOR EDUCATION POLICY

II. Identifying Education Policy Principles (EPPs)

In the past several decades of education in the United States, certain fundamental precepts can be distinguished by scrutinizing the effects, positive and negative, of past education reform and policy. This chapter identifies these in the form of five Education Policy Principles (EPPs).[2] These EPPs diagnose the core areas that future reforms, at no particular level of government,[3] must address to rectify: the inequity that predates accountability reforms, the issue of high stakes, and the unintentional byproducts of high-stakes accountability. This chapter will first summarize the political background for equity and excellence movements in education, providing a historical framework with implications for future policy. Following this discussion, the identified EPPs will be evaluated and defended by references to research and existing literature. Consequently, the following sections review evidence establishing the validity of EPPs 1, 2 and 3, on equal standards, need versus merit, and higher-order learning, respectively.

Evidence justifying EPP 4 and EPP 5, which discuss stakes and multiple indicators for accountability (respectively), will be presented in depth in the following chapter.[4] These chapters will examine persisting effects of NCLB-era reforms to ultimately identify the dominant issue in the ongoing system of accountability: high stakes. At one point underpinning federal-level accountability under NCLB, these high stakes now define accountability at the *state* level – the primary change enacted as NCLB was replaced by ESSA.[5]

In the following chapters, ability to observe the five EPPs will become the basis on which

---

[2] See Figure 2 "Education Policy Principles (EPPs)" on the following page.
[3] Accordingly, the Policy Recommendations (PRs) addressing these EPPs will occur at various levels of government – federal, state, or local – in the following chapters of this paper.
[4] See: III.i: *EPP 4: High Stakes are Detrimental in Systems of Accountability*, and III.ii: *EPP 5: Campbell's Law*
[5] More on this in Federal Chapter IV, section i

Policy Recommendations (PRs) at the federal, state, and local levels are argued. Thus, these compiled principles function as axioms in strides toward equity at the intersection of federal, state, and local accountability. The EPPs, then, *diagnose* the issues, while the PRs *prescribe* solutions.

Figure 2: Education Policy Principles (EPPs)

The table below lists five principles for education policy. These principles (EPPs) have been identified from an analysis of the effects of past policies, which included examining trends in student performance and reviewing the current literature on education research. These principles should, in theory, be addressed as future reforms attempt to mitigate inequities in the current American education system. This chapter will defend and explore each EPP individually, in an in-depth analysis. Note: Principles (EPPs) 4 and 5 will be addressed independently in chapter III, "Stakes and Accountability."

Principle 1	Merit-based policies undermine equity by supplanting need-based funding and support.
Principle 2	Assessments should facilitate development of higher order learning and 21st century skills.
Principle 3	Performance measurement and associated standards for growth must remain equitable among all populations of students to ensure equal opportunity for academic achievement.
Principle 4	High-stakes accountability serves to hinder rather than incentivize educators in systems of accountability.
Principle 5	A system with multiple indicators of student performance produces better, more equitable data.

i. Historical and Political Background

In 1990, seven years after *A Nation at Risk* spurred a widespread perception that the country faced a crisis of "mediocrity" and declining performance in United States schools, Sandia National Laboratories, consisting of two U.S. Department of Energy research and development laboratories, was commissioned by Department of Energy Secretary Admiral James Watkins to produce a report aggregating data on the state of education in the United States. The Sandia report, titled "Perspectives on Education in America," released in published form in the

summer 1993 issue of *The Journal of Education Research.* The report quickly rose to controversy when its data indicated a state of education in contradiction to the notion of steep performance decline on which, at that point in time, George W. Bush centered his presidency. The Sandia report, which used data from SAT and NAEP performance in addition to the International Assessment of Educational Progress (IAEP), concluded that student performance data was actually displaying "steady or slightly improving trends" over the ten year period (Carson, Heulskamp and Woodall 1993). The position taken in the report's conclusion – that though there were gaps between the performance of white and minority students, all demographic groups demonstrated steadily inclining performance – debunked the crisis concept on which education reform in the 1990's gained political support (Carson, Heulskamp and Woodall 1993). Those findings remain heavily disputed due to apparent flaws in data usage and aggregation: the report referenced only the 1988 results on the IAEP, omitted review of relevant state standardized tests, and failed to provide substantiating evidence to back points or citations to back the data (Stedman, 1994).

Studies conducted by other researchers, however, align with claims made in the Sandia report, suggesting at least some level of validity. In 1985, Gregory Anrig, then president of the Educational Testing Service, pointed to improving performance among black students[6] and closure of the achievement gap[7] as evidence of "progress toward educational equity." Following calls for reform that echoed claims in *A Nation at Risk,* Anrig warned that, "Excellence must not

[6] "The [NAEP] reported that black students at the three age levels tested (9, 13, and 17) showed steady and consistent improvement in reading from 1970 to 1980. Mathematics scores for all students declined overall from 1972 to 1978, but scores improved for black 9-year-olds, remained stable for black 13-year-olds, and declined no more than those of white students for black 17-year-olds" (Anrig , 1985, p. 624).
[7] "In the eight-year period between 1976 and 1984, mean score differences on the SAT between all students and black students were reduced by 13% on the verbal portion and by 18% on the mathematical portion of this nationally standardized college admissions test" (Anrig , 1985, p. 624).

become the new code word for a retreat from equity, just when the struggles of recent years are beginning to pay dividends" (Anrig, 1985, p. 624). Jaekyung Lee, in the Graduate School of Education at the State University of New York at Buffalo, produced research in 2002 corroborating Anrig's comments on the achievement gap: "Racial and ethnic achievement gaps narrowed substantially in the 1970s and 1980s when Blacks and Hispanics improved their reading and mathematics achievement much more than did their White counterparts" (Lee, 2004, p. 52). Lee noted, too, that these trends reversed in the 1990s as the gap again widened[8] – potentially confirming Anrig's fear that progress toward educational equality would be set back.

Regardless of the legitimacy underpinning a "crisis" necessitating the implementation of broad education reform from the top – at the federal level – down to the local level, it has since become clear that the disparity in student performance between racial and income groups is in its own right a national crisis. Since the Coleman Report in 1965, the gap in achievement has only marginally lowered or stayed the same, and in some performance areas, such as reading results in the Midwest, the gap as widened even further, according to analysis by Education Next.[9] If the crisis was missing in the Reagan and Bush -eras, it certainly has since grown to significant existence, not only worthy of, but necessitating, political attention.

As we approach the two-decade mark since the initial implementation of No Child Left Behind, it has become increasingly clear that efforts to close the achievement gap are not demonstrating the educational equality that they originally prescribed to achieve (Guisbond, et. al, 2013, p. 7). Additionally, the overall improvement in student performance is increasingly delegitimized by the charge that standardized tests measure only rote, second-order skills

[8] The policy implications of these fluctuating trends in the achievement gap will be discussed in-depth in the following chapters: See Chapter IV, "Policy Recommendations at the Federal Level."
[9] "Achievement Gap Between White and Black Students Still Gaping." *Education Next.* (See: https://www.usnews.com/news/blogs/data-mine/2016/01/13/achievement-gap-between-white-and-black-students-still-gaping)

(Darling-Hammond 2000, 3). NCLB and the standards-based reforms that have since followed have not, however, been for naught. We have developed decades of additional useful data, particularly data that indicates the state of performance for previously ignored demographics in education, which allows us to further consider and evaluate the soundness of methods to equalize performance and educational opportunity among all students. The lack of progress toward equal achievement in the past two decades, however, has indicated the necessity for the adoption of a new perspective on reform.

In the following sections, I discuss principles (EPPs 1, 4, and 5) to underpin future considerations on education reform as we move forward in new attempts to close the achievement gap. These EPPs combine with the previously discussed EPP 2 and EPP 3 to constitute five principles for policy in education, as we move toward a different, more effective means for achievement in all demographic communities:

ii. EPP 1, EPP2, EPP3: Discussion of the Principles

EPP 1: *Departing from Merit: Mistakes of the 'Excellence Era'*

The 'Excellence Era' in education reform began in the 1980's, shortly following NCEE's release of *A Nation at Risk*. A statement made by Indiana Senator Richard Lugar at the (year) National Forum on Excellence in Education captures the political zeitgeist in the sphere of education: "We stand on our heads to stimulate athletic achievement, and our children achieve in spectacular fashion. Underpaid or even unpaid coaches bask in community esteem. But when it comes to academic achievement, we have still not caught the spirit. This is the crisis of American education . . . We have been a nation at risk because we have not cared enough about learning." (Senator Lugar's comments reflected the political focal shift from need and equity, to

merit and achievement– the excellence movement in education reform. Arguments for this shift similarly centered on the perception of a "crisis of American education." The movement, too, "departed substantially from the previously dominant equity-in-education paradigm, stimulating important departures in education politics and policymaking," and continues to "set the tone for education policy up to the present day" (Rhodes, 40-41).

The focal shift toward merit stimulated government control of standards and curricula, which in turn began extensive standardized testing.[10] Merit, by the nature of the system, necessitates means by which "achievement" can be demonstrated, and when this demonstrable performance follows the student, "achievement" can have important implications for the student's future opportunity: "High school exit exams, standardized and quantified student and teacher assessment, a strengthened back-to-basics core curriculum – all aspects of meritocracy – function if not to maintain the status quo, to at least regulate tightly incremental social change." At this point, student performance develops high stakes. Other elements within and implications of high-stakes accountability, a primary consequence of a meritocracy, necessitate independent attention, and will be further discussed in EPP 4.[11]

Importantly, the shift from need to merit disadvantages students whose ability to demonstrate merit is marginalized at the outset by low income or minority race (Robenstine 1997). The meritocratic education system, appealing in many respects to the American ideal that success directly aligns with work ethic, assumes that those who "fail" by performing below

[10] "Consequently, state governments, though not yet the federal government, began to engage in more vigorous efforts to influence standards, curricula, testing, and teacher quality" (Rhodes, 40).
[11] See Chapter II, section i: "EPP 4: High Stakes are Detrimental in Systems of Accountability"

average do so by lack of ability and ethic, rather than by lack of opportunity.[12] Clark Robenstine, education researcher at the University of South Alabama, argues that acceptance of the merit ideology is justifiable only by first ensuring the existence of the equally American ideology of *equality*:

> "Thus far, contemporary schooling has been generally unable to interrupt the deterministic relationship between socioeconomic status and schooling outcomes, and education reform policies which maintain the status quo do little to alter this. The reason continues to be a blind adherence to meritocracy without providing the prior necessity of equal educational opportunity." (Robenstine 1997, 117)

The enduring meritocracy of the American education system has, for decades, mistakenly assumed the equality prerequisite.

Research, however, has consistently confirmed that equal opportunity has not existed, and still does not exist, in American education. Federal studies[13] have indicated that income and poverty levels have led to "very large differences in the acquisition of skills invaluable for school success long before the children ever enter a schoolhouse" (Orfield and Lee 2005, 5). Inequalities in socioeconomic-level are subsequently reflected in results on high-stakes standardized tests.[14] In addition to inequities stemming from socioeconomic differences, race remains an issue in US schools: "Levels of segregation for black and Latino students have been steadily increasing since

---

[12] "Those who fail to achieve socially and economically do so "obviously" because they first failed to work hard, study, learn their lessons, pay attention, and do what the teacher told them - they lacked individual ability and motivation. Thus, social and economic failure is justified. By virtue of its centrality to American ideology, meritocracy ensures perpetuation of that dominant ideology" (Robenstine, 116).

[13] Rathbun, A., West, J., and Germino Hausken, E. (2004). *From kindergarten through third grade: Children's beginning school experiences.* Washington, DC: National Center for Education Statistics. (Discussed in Orfield and Lee, 5)

[14] Orfield and Lee point to a study* of metro Boston to confirm this: In the metro region, "97 percent of the schools with less than a tenth white students face concentrated poverty compared to 1 percent of the schools with less than a tenth minority students. These differences were strongly related to the results on the high stakes MCAS state examinations" (Orfield, 6; *Study from: Lee, C. (2004). *Racial segregation and educational outcomes in metropolitan Boston.* Cambridge: The Civil Rights Project at Harvard University.).

the 1980s . . . Achievement scores are strongly linked to school racial composition and so is the presence of highly qualified and experienced teachers" (Orfield and Lee 2005, 5). This correlation between persisting inequities and student performance confirm problematic elements of a merit-based system. In the face of the American achievement gap, meritocracy is thus shown to perpetuate – and perhaps strengthen – existing social and academic inequities.

The implications of merit in systems of accountability will be further explored in Federal PR 1, which will examine how positive trends in narrowing the achievement gap that first emerged under ESEA (need-based aid under Title I) began to reverse with excellence era reforms. Finally, State PR 2 will recommend a returning shift from the merit-based accountability system, to one that again accounts for need.

EPP 2: *Facilitating Higher-Order Learning and 21st Century Skills*

As discontent following NCLB spurred debate on the growing role of assessments in the classroom, discussion began to center on learning. In particular, studies found that the rote nature of standardized tests combined with the high-stakes to "narrow the curriculum, pushing instruction toward lower order cognitive skills" (Darling-Hammond 2000, 3). Simultaneously, in the transforming economic landscape of the twenty first century, "business operations have become so globalized that core business competencies place greater emphasis on knowledge, mobility, and collaboration" (Dunning 2000, as cited in Kai Wah Chu, et al. 2016, 17). These increasing demands of the workplace have led to calls for public education to assure student engagement in higher-order learning and development of critical thinking skills – 'twenty first century skills.'

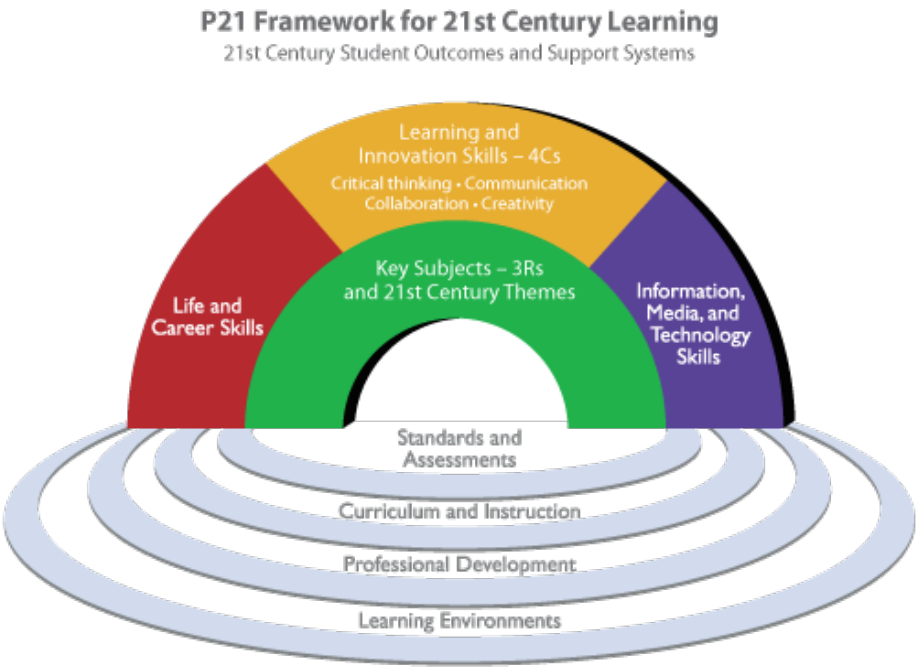


Figure 3: Partnership for 21st Century Skills 'Framework for 21st Century Learning'
Above is a visual representation of P21's framework for twenty first century learning, with twenty first century skills and themes at the center, with other skills and educational institutions in the surrounding circle of the framework. (Adapted from Partnership for 21st Century Skills, http://www.p21.org/)

Critics of the label argue that it is misleading, because the skills it encompasses, "Knowing how to think critically, analytically, and creatively," have been valuable throughout history, and are thus not "unique to the 21st century" (Silva 2009, 631). However, a long line of research points to the growing *necessity* of critical cognitive skills, as the needs of the workplace are rapidly evolving. Skills that were unnecessary several decades ago, the ability to read, for example, are now baseline competencies for employment, even in the service sector[15] (Murnane and Levy 2005, 16). There are logical reasons underlying the rising competency levels: the Internet and

[15] "25 years ago, auto mechanics did not have to read to learn their jobs – they could learn by watching other mechanics. But the evolution of automobile electronics has transformed many visible, mechanical components into opaque electronic modules. As a result, a mechanic can no longer function without the ability to read and to work with computerized testing equipment" (Murnane and Levy 2005, 16).

information –era has created an "exponential growth" of information, such that content may seemingly expire if not continually updated to present demands (OECD 2004, Pedró 2006; as cited in Kai Wah Chu, et al. 2016, 18). Thus, everyone must be "prepared for and convinced of the need to be lifelong learners to keep pace with the evolution of technology (Medel-Añonuevo et al. 2001; as cited in Kai Wah Chu, et al. 2016, 18).

Further, in a report for the Industrial Performance Center at Massachusetts Institute of Technology, Frank Levy and Richard Murnane found that advancing technology increasingly equips computers to execute tasks formerly completed by humans, causing a growing displacement of available jobs (Murnane and Levy 2005, 23-24). Murnane and Levy argue that, because all human work consists of cognitive tasks to process information, and because computers execute rules, "a computer can substitute for a human in processing information when two conditions are present: The information to be processed can be represented in a form that is suitable for use by a computer [and] The processing itself can be expressed in a series of rules" (Murnane and Levy 2005, 3). As a greater number of lower-level jobs are replaced by computers, those jobs become less available to humans, which heightens the urgency to develop skills necessary for more cognitively demanding employment opportunities. Thus, the demand for public education to address these skills with students becomes increasingly imperative.

These demands are plentiful – numerous studies indicate that the U.S. workforce in general (Manyika et al. 2011; as cited in Burrus et al. 2013, 1), including large portions of college graduates (36% for 4-year graduates in 2011) is deficient in the skills required by employers (Casner-Lotto & Barrington 2006; as cited in Burrus et al. 2013, 1). Consequently, multiple researchers and public and private interest organizations have conducted studies to identify exactly which skills are relevant for the twenty first century learning demands. An analysis of the Occupational Information Network (O*NET) for the Educational Testing Service in 2013

synthesized three frameworks[16] for twenty first century skills to identify competencies important for employment:

> "A comparison of [the ranking of identified competency components] with previous 21st century competencies frameworks suggested that 5 competencies stand out as important for most occupations: problem solving (e.g., complex problem solving), fluid intelligence (e.g., category flexibility), teamwork (e.g., cooperation), achievement/innovation (e.g., persistence), and communication skills (e.g., oral expression)" (Burrus et al. 2013, i).

Further, the report found that, "4 of these 5 competencies were strongly related to wages" ((Burrus et al. 2013, i), indicating that the skills developed in K-12 education – which inform competency for employment – have larger implications for the wage gap.

As this report indicates, the demands for higher-order critical thinking skills are widespread and have important ramifications for students once they enter the workforce. Failing to address the deficiency of higher-order learning in education will allow the trend cited by Burrus et al. to continue; the achievement gap could quickly become another generation's wage gap. Further, these concerns are not limited to the employment sector: David Conley writes in *College Knowledge* (2005) that institutions of higher education value "habits of mind," which includes critical thinking skills and analytical abilities, more so than mastery of content (Conley 2005; as cited by Stanford Research Network). To address the growing demands that students develop twenty first century skills, the PR at the local level proposes policy to implement a statewide system of Performance-based Assessments. These assessments consist of project-based performance tasks, each of which requires students to engage in critical analysis, to measure student achievement.

[16] The three frameworks analyzed in the datasets were: Finegold and Notabartolo (2008), which conducted a literature review; Assessing and Teaching of 21st Century Skills (ATC21S), which employed a group of workforce experts (Binkley et al., 2010); and Performance for 21st Century Learning (P21), which employed panels of experts to define a framework for twenty first century skills. (Burrus et al. 2013)

EPP 3: *Equity Necessitates Accountability Standards*

More than six decades since Brown v. Board of Education ended the legal practice of segregation in schools, inequality still flagrantly marks the American education system. In 2013, the average black twelfth grade student scored only in the 19th percentile of the white distribution for math, and only in the 22nd percentile of the white distribution for reading.[17] These evident inequities necessitate the use of accountability standards, as a means to measure and monitor progress toward equity. 'Standards' in education refer to general or specific identified objectives for academic achievement in certain subject areas, such as math or English language arts, for *all* students. 'Accountability' refers to the practice of holding institutions of education accountable for progress toward these standards, for *all* students. The responsibility for inclusiveness inherent to these educational practices makes accountability standards a fundamental element of progress toward equity in education.

Standards-based reforms – reforms centered on accountability – emerged in the 1990s under the Clinton-era Improving America's Schools Act (IASA), and later strengthened to become the cornerstone of policy under the Bush administration's passage of NCLB. The reform was well intentioned: As written in the legislation, the goal of NCLB was "to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and State academic assessments."[18] Despite its intentions, NCLB was, and still is, widely criticized for its effects on marginalized populations: "NCLB has severely damaged educational quality and

[17] "In math, the size of the gap has fallen nationally by 0.2 standard deviations, but that still leaves the average black 12th-grade student at only the 19th percentile of the white distribution. In reading, the achievement gap has improved slightly more than in math (0.3 standard deviations), but after a half century, the average black student scores at just the 22nd percentile of the white distribution." (Hanushek and Peterson, "What Matters for Student Achievement." *Education Next.* Spring 2016. Podcast.)
[18] See 20 *U.S. Code* (*U.S.C.*) 6302 § 1001.

equity, with its narrowing and limiting effects falling most severely on the poor . . . [It] failed to significantly increase average academic performance and significantly narrow achievement gaps" (Guisbond, et. al, 2013, p. 7). However, the notion of standards and assessments that defined the legislation are not, in themselves, responsible for NCLB's negative ramifications.[19]

Rather than to induce "excellence," standards can, and must, be used to attain and maintain equity in education. Paul Gagnon, researcher in the School of Education at Boston University, argued for standards as a means toward equity during the onset of standards-based reform: "We already know what excellence is; we have high academic standards in some of our schools, public and private. The need is equality: to apply those standards to the schools that all the rest of our children find themselves in" (Gagnon, 1994, 2). Gagnon's comments highlight the potential of academic standards – both for use, toward equity and quality of education, and misuse, toward high-stakes performance in the name of "excellence." When employed to hold schools accountable for progress toward equity, standards give rise to quality of teaching and learning for *all* students (Darling-Hammond 2000). Linda Darling-Hammond, President and CEO of the Learning Policy Institute and professor in the Graduate School of Education at Stanford University, argues that accountability is a process, and standards – in their creation and assessment – are only a single element of the process. In addition to developing "criterion-referenced" learning standards, she recommends using, "standards-based performance assessments of student achievement as diagnostics to guide improved teaching and needed supports, not as the basis for punishing students" (Darling-Hammond 2000, 5).[20] In usage for diagnostic purposes, standards simultaneously serve all students in indicating areas for

[19] Rather, *high-stakes* tied to the standards led to much of the residual damage: See Chapter III, section i, "High-Stakes are Detrimental in Systems of Accountability."

[20] Darling-Hammond makes these recommendations with a set of other proposed policies – such as using standards to hire and support teachers, and assuring that curricula is responsive to the standards – to ensure that standards support learning (Darling-Hammond 2000, 5).

improvement, and facilitate equity in providing means for identifying inequities when and where they occur.

Standards, by this definition, necessitate accountability. The immediate question, then, turns to *how* schools should be held accountable, and this discussion generally turns toward assessment. These assessments, when attached to high-stakes, can "narrow the curriculum, pushing instruction toward lower order cognitive skills, and can distort scores" (Darling-Hammond 2000, 3). The evidence of this effect following NCLB drew criticism of a perceived testing-based nature of the reform.[21] However, assessment in some form[22] is necessary to measure the standards in place for ensuring equity in the classroom. Anrig argues for the utility of unbiased assessments in eliminating educational inequities:

> "When minority students perform less well than non-minority students on a given test, people too often assume automatically that the test must be biased against minority students. When the test is not biased, however, to attack its use because of unfavorable results is to attack a potential force for improving those results. We do not cure a virus by throwing away the thermometer that alerts us to the existence of a fever. So it is with unequal educational opportunities." (Anrig 1985, 624)

Despite the pressure testing has introduced to the classroom (Bussert-Webb 1999, Moses and Nanna 2007), some teachers agree with this assessment. Farida Mama, a fifth-grade math teacher at a public charter in Boston and a policy fellow at TeachPlus, met with President Obama during his administration to discuss testing. She is quoted by Slate as pro-testing: "I believe that for equity and civil rights issues, it's essential that the government holds all schools and districts accountable for student achievement." She also acknowledged that, "In our efforts to get data we sometimes forget how quickly the tests can add up." On the opposite side stand teachers who

[21] See Chapter III, section i, "High-Stakes are Detrimental in Systems of Accountability."
[22] Assessment for purposes of equity does not *necessitate* statewide standardized testing among every student. Chapter V, "Policy Recommendations at the State Level," proposes policies for maintaining necessary performance data, eliminating the testing of every student with statewide standardized assessments, and exploring other non-standardized means for assessing every student.

staunchly oppose testing, though many reasons for opposition can be traced back to the stakes attached to the assessment, rather than the assessment itself.[23]

The argument for accountability standards is typically centered on its potential to improve overall student achievement, rather than its potential to facilitate a more equitable education system. And yet, equality is an inseparable component of this argument: Introduction of the first standards in the 1960s provided the data that alerted us to the issue of the achievement gap, and a platform for implementing new policy in order to address that gap. The objective now, then, becomes creating a system of accountability in which assessments inform the instruction, rather than instruction informing assessment. Federal PRs 1 and 2, and State PRs 1 and 2, propose policies to achieve such a system – a *new* accountability.

While this chapter presented, discussed, and defended the first three EPPs, the next chapter of this paper will explore the final two of the five EPPs (EPPs 4 and 5) which both address principles gleaned from mistakes of past and current systems of accountability. This following chapter – "Stakes and Accountability" – will center on high-stakes accountability and methods performance measurement that can potentially add stakes to particular indicators of achievement.

[23] See Chapter III, section i, "High-Stakes are Detrimental in Systems of Accountability."

III. Stakes and Accountability

Chapter II identified and introduced five principles for education policy (EPPs), each of which future education reforms must address in order to prevent negative outcomes similar to those unintentionally resulting from previous reforms. EPPs 1, 2 and 3 – regarding need-based versus merit-based policy, twenty first century skills, and the role standards play in enabling equity, respectively – were each discussed and defended in depth. The discussion of EPP 3, which concluded the previous chapter, ended with the suggestion of a *new* accountability. Chapter III will explore elements of standards-based education policies that have produced unintended consequences, such as corruptible data and testing-based instruction, in previous systems of accountability. This chapter discusses EPPs 4 and 5, on high-stakes accountability standards and reliance on multiple indicators of student performance (Campbell's Law), respectively.

i. EPP 4: High Stakes are Detrimental in Systems of Accountability

No Child Left Behind: The Flawed Argument for High-Stakes

At a time in which education reform initiatives[24] seem to have become the crusade of party ideologies on both sides of the political spectrum, high stakes testing[25] has fixed itself at the critical intersection of education, standards measurement, and the law. Many states still sustain significant effects from the punitive sanctions first significantly introduced with Bush's NCLB, the

[24] The current administration's latest "crusade," under newly confirmed U.S. Department of Education Secretary Betsy DeVos, is school choice. The Choices in Education Act of 2017 (H.R. 610), introduced in the House on Jan. 23, 2017, repeals ESEA and "establishes an education voucher program, through which each state shall distribute block grant funds among local educational agencies (LEAs) based on the number of eligible children within each LEA's geographical area." See: Choices in Education Act of 2017, H.R. 610, 115th Cong. (2017). (https://www.congress.gov/bill/115th-congress/house-bill/610)

[25] A "high stakes" test is any form of testing in which student performance has significant consequences for students, teachers, or principals and administrator.

impact of the sanctions manifesting in high-stakes statewide assessments. At this point, however, high-stakes testing as a means for accountability – among students, teachers, and principals alike – is not a new phenomenon. Accordingly, there now exist long lines of research assessing the validity, equity and effectiveness of high-stakes tests themselves. The findings increasingly point to negative effects of high-stakes: While stakes were originally intended to incentivize and thus stimulate student achievement, creating consequences for student performance has produced unintended, arguably harmful, ramifications at every level (Bussert-Webb 1999, Marker 2001, Koza 2002, Berliner and Nichols 2005, Moses and Nanna 2007, Mintrop and Sunderman 2009, FairTest 2010, Phelps 2011).

Preceding and following Bush's NCLB, right and left wing lawmakers lauded standards as the ultimate tool to promote both equity and achievement in the classroom.[26] EPP 3 confirms the validity of this perspective – standards enable equity. Issues arise, however, when policies tie stakes to accountability standards. While contemporary standards-based reforms began with the Clinton administration's 1994 passage of Improving America's Schools Act (IASA), stakes did not have a significant presence until NCLB. The punitive nature of NCLB's sanctions created a culture that relied on fear to influence teacher success: "What is strictly stipulated … is the staged progression of underperforming units through a set of increasingly severe sanctions based on meeting performance quotas for specific demographic groups: from identification and publication of 'school improvement' status (a kind of public shaming with potentially far reaching market consequences); mounting loss or organizational autonomy through required external

[26] There is issue with the argument that high stakes initiatives such as NCLB provide needed insight into achievement gaps. "Achievement gaps were well known prior to NCLB. The disaggregation of NAEP test results has provided clear documentation of achievement gaps for many decades. What NCLB and related policies added was a set of punitive interventions, not a guiding knowledge of the gaps and not a set of strategies and resources to close the gaps" (Welner and Mathis, 246).

intervention and severe contracting; and finally termination through re-organization or take-over of the organization" (Mintrop and Sunderman 2009, 63-64).

At the core of the argument for maintaining stakes on statewide assessments is the performance data produced by frequent assessment, and the subsequent ability to set adequate standards in response to that data. As discussed in EPP 3, testing provides a platform from which to set and measure standards for student achievement. For example, supporters of high school exit exams – the stakes, of course, being graduation – argue that the practice motivates preparedness for college-level academic work. Thomas Vukovich, an associate provost at the University of Akron, believes Ohio's graduation exam is responsible for ensuring the preparation of incoming students: "I think it [Ohio's exam] is sending a message. It is saying, 'If you want to come to college, you have to prepare for it and be more serious about it in high school'" (Schmidt 2000; as cited in Moses and Nanna 2007, p. 58). However, critical analyses indicate that proliferation of high-stakes tests does not reflect significant improvements in performance, especially compared to other education systems. The Board on Testing Assessment determined that a system of incentivized testing does not significantly improve achievement in the United States to the levels of other countries performing higher in education.[27] While low stakes tests, which were less likely to reflect inflated results, showed very little and, in many cases, no effect at all on achievement, higher stakes incentivized testing showed only minor effects.[28] The board additionally determined that, particularly with testing used for graduation requirements, the stakes can serve to degrade achievement rather than promote it: "The evidence we have

[27] ""When evaluated using relevant low-stakes tests, which are less likely to be inflated by the incentives themselves, the overall effects on achievement tend to be small and are effectively zero for a number of programs. Even when evaluated using the tests attached to the incentives, a number of programs show only small effects" (Hout and Elliott 2011, p. 237).

[28] The only incentivized measures showing significant effects were those incorporated within NCLB, which still did not approach levels of achievement intended by the testing initiatives. (Hout and Elliott, "Incentives and Test-Based Accountability in Eduation," 237).

reviewed suggests that high school exit exam programs, as currently implemented in the United States, decrease the rate of high school graduation without increasing achievement" (Hout and Elliott, 238).[29]

Furthermore, supporters argue that the administrative utility of standardized testing – in its production of relatively simple performance data – validates the usage of stakes. In addition to providing a means for benchmarks at the progression of grade levels, standardized tests such as the SAT and the ACT provide a relatively stable means of assessment for college admissions. Confronted with large numbers of applications that require a quick turnaround on decisions and some measure for accurate, objective judgment, college admission offices look to these tests to provide a cost-effective and efficient means for evaluating large amounts of information. However, the ease of processing is accompanied by an increasing reliance on the SAT and ACT as standards for measurement of a student's potential for success at the college level. In an investigation of admissions standards at several prominently known graduate schools, standardized test scores appeared to hold significantly more weight than GPA: "A high standardized test score coupled with a high or moderate GPA was associated with the highest rate of acceptance," while, interestingly, "a low GPA coupled with a high standardized test score had a higher acceptance rate than a high GPA and low standardized test score" (Millimet & Flume 1982; as cited in Moses and Nanna 2007, 60). Stakes for post-secondary educational opportunity, these studies indicate, have shifted from actual academic performance on a typical class day (GPA) toward performance on a single hours-long standardized assessment.

[29] They suggest that a rewards-based approach, rather than the current punitive model, may be more productive: "The best available estimate [for effect on achievement] suggests a decrease of two percentage points when averaged over the population. In contrast, several experiments with providing incentives for graduation in the form of rewards, while keeping graduation standards constant, suggest that such incentives might be used to increase high school completion" (Hout and Elliott, "Incentives and Test-Based Accountability in Eduation," 237).

Standardization and Teaching to the Test

A growing body of evidence indicates that stakes placed on teachers and principals, by assessing their students' performance on high-stakes tests, is creating an unfortunate phenomenon: teachers feel compelled to "teach to the test," and principals, studies find, feel compelled to support and encourage class time spent on test preparation (Phelps 2011). Rather than functioning at the service of the instruction for teachers, high stakes have generated teacher instruction that functions to serve the test (Phelps 2011). Arthur Coleman, Deputy Assistant Secretary of the U.S. Department of Education's Office for Civil Rights from 1997 to Jan. 2000, and a proponent of appropriate usage of high stakes, unintentionally illustrates the circumstances that produce test-centered instruction. Teaching the content tested, he correctly argues, is necessary to ensure that the testing itself is fair, such that students receive fair and equal opportunity to display merit and achievement via performance:

> "In the context of a high-stakes test that is designed to measure what a student has learned, this means that the material on the test must be aligned with the curriculum and instruction of the student being tested. In other words, the material tested on an achievement test to which high stakes consequences are attached should be the same material that the student has been taught, providing him with a fair opportunity to learn the material that is being tested. [8] If the content of instruction and the teaching methodologies provide the student the opportunity to learn the material being tested, the test is more likely to be valid, reliable and fair" (Coleman 1998, 106).

Coleman, however, fails to address the consequences of such a system in the context of an achievement test that measures only rote skills and basic content, and the further implications for low performing classrooms – with primarily low-income and minority demographics – that have the most at stake. The harm is magnified when schools buckle under the pressure of high stakes testing, ultimately eliminating entire subjects that will not be tested: "Facing high-stakes test-based accountability under NCLB and state laws, schools narrow curriculum by reducing or

dropping untested subjects. The law promotes teaching to the mostly multiple-choice state tests, focusing one-sidedly on rote skills and ignoring higher-level thinking," and worse, "The impact is greater in schools that serve low-income youth, particularly students of color." (Neill 2009; as cited in FairTest, "Racial Justice and Standardized Educational Testing").

Kathy Bussert-Webb, a teacher-researcher in a 98% Mexican American high school in South Texas, wrote about the harm of teaching to the test. Her findings are informed by student feedback, in addition to her realization that her own teaching styles had conformed to the objectives emphasized by the Texas Achievement of Academic Skills (TAAS) graduation requirement test. After being told by freshman that their reason for dropping out was expectance of failure on the TAAS and unwillingness to continue retaking the exam, Bussart-Webb believes that, "a positive relationship may exist between a standardized test focus and our high school dropout rate." She continues, "Many of my students said they didn't care about TAAS and that they were tired of it . . .Many said they loved reading and writing until fifth grade, when their teachers started emphasizing TAAS. They said this sole focus made them feel pressured and nervous about school and that their teachers gradually made classes less enjoyable. They had fewer field trips, art activities, games and interesting stories" (Bussert-Webb 1999, 582).

Implications for Racial & Income Equality

The policies creating stakes are admittedly well intentioned; improving educational achievement is righteous cause. As established previously, supporters of high stakes tests argue that the standardization of testing maintains equity and objectivity in a system that can become subjectively corrupted when factors such as race, gender or income-level have room to effect a student's educational opportunity – and thus that this equitable foundation provides a valid basis

for punitive sanctioning in order to manufacture better results and ultimately improve the system. This, supposedly, was the intention underlying the proliferation of high stakes testing. However, overwhelming amounts of data and evidence suggest that, contrary to the intended effect, the stakes tied to testing have, if anything else, buttressed the already strongly intact systemic inequalities within education (Jones and Whitford 2002, Neill 2009, Madaus and Russell 2010, Kearns 2011, McDermott 2011).

Supporters of high stakes often dispute their impact on marginalized demographic groups. Coleman, for instance, argues that despite the large variations in performance between middle-class Caucasian students and low-income minority students, high-stakes tests constitute a valid comparison of student performance: "Any conclusion, for instance, that the federal civil rights laws require the performance of different racial groups to be equal is fallacious. The requirement is that each child have an equal *opportunity* to succeed; it does not require equal results" (Coleman 1998, 100, note 63). However, numerous studies have found that that low-income minority students disproportionately fail (what Coleman claims are fair and equal) assessments, with stakes such as graduation and post-secondary education that limit future opportunities (Natriello and Pallas 1999; Nichols, Glass, and Berliner 2005; as cited in Moses and Nanna 2007), These inequities garner additional concern in the context of "stereotype threat" – a term used by Claude Steele, a professor of social psychology and president-elect of the Western Psychological Association, to describe the potential for minority race and gender status to psychologically impact student performance on standardized tests. The seven-year study found that stakes compounded the impact of stereotype threat.[30]

[30] "As Claude Steele and his colleagues have demonstrated, 'stereotype threat' increases the likelihood that students of color will have inaccurately low scores. Stereotype threat means that students who are aware of racial and gender stereotypes about their group's intellectual ability score lower on standardized tests perceived to measure academic aptitude. In effect, the use of high-stakes testing in an overall

According to analysis by FairTest, the National Center for Fair & Open Testing, the populations suffering the most – via restricted opportunity – from a system relying heavily on high stakes tests are racial minorities, and particularly those who come from low-income households, for whom the harm is compounded. This raises concerns about equity particularly in regard to findings that performance can further limit educational opportunity: An analysis by the Advancement Project et al. found that low performers may face discriminatory action. The paper reports that schools "at times suspend, expel, 'counsel out' or otherwise remove students with low scores in order to boost school results and escape test-based sanctions mandated by [NCLB], at great cost to the youth and ultimately society" (Advancement Project et al. 2010; as cited in FairTest). The stakes associated with student performance thus appear to lead to prioritization of high performing students over low performing students, a problematic discovery considering the disproportionate failure of low-income and minority demographics with limited opportunities. Jennifer Booher-Jennings, in a 2005 study of the Texas accountability system for Columbia University found that while some students were ignored because they were likely to fail, others were placed in special education to avoid lowering the schools' overall performance averages. Students who received the most attention, Booher-Jennings also found, were those on the border of passing and failing. Additionally, the attention was not always beneficial to the students: in some instances, students were pulled from extra-curricular classes such as music and gym, in order to devote additional time to test preparation (Booher-Jennings 2005).

The numerous instances of unintentional, yet harmful, effects that high stakes tests have had on students – and particularly students belonging to demographic populations marginalized by income, race and/or gender – indicate the necessity to adopt alternative methods of

---

environment of racial inequality perpetuates that inequality through the emotional and psychological power of the tests over the test-takers" (Claude Steele 1995; as cited in FairTest; *see: "Stereotypes Lower Test Scores"*).

assessment in future systems of accountability at the state level. Federal PRs 1 and 2 propose policies that alter guidelines at the federal level to allow states to lower the stakes tied to statewide assessments. State PR 1 subsequently recommends a system of measurement that lowers the stakes of student performance on standardized assessments, thus addressing the consequences of high stakes discussed in this EPP.

ii. EPP 4: Campbell's Law: The Necessity of Multiple Indicators

A common critique of high-stakes testing is the anxiety the stakes foster from the student perspective, functioning – instead of motivating and incentivizing the student to take active interest in their educational achievement – to lead students to give up on what becomes the methodical and seemingly impossible task (Bussert-Webb 1999). However, as Booher-Jennings found, stakes on educators and schools can magnify negative consequences for students in the attention directed to various performance demographics (Booher-Jennings 2005). The impact of stakes on teachers, such as salary and security in employment, and schools, such as funding, thus illuminates an additional area for concern.[31] The result of high stakes decisions affecting educators is what Nichols and Berliner have called "single-indicator corruptibility," a concept rooted in Campbell's law: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort

[31] High stakes decisions "affect teachers and principals when scores are used to determine merit pay or potential dismissal. Still others affect schools, as when schools are awarded recognition or extra funds when scores increase or are put into intervention status or threatened with loss of registration when scores are low. Some policies take into account differences in the initial performance of students and in the many non-school factors that can affect achievement. Some do not, holding schools to similar standards despite dissimilar student populations and resources" (Darling-Hammond 2000, 2).

and corrupt the social processes it is intended to monitor" (Campbell; as cited in Berliner and Nichols 2005, i).

Because standardized tests seem to increasingly serve at the cornerstone of important decisions regarding funding and allocation of resources, the stakes remain high for schools in numerous states across the nation. In August 2016, for example, Texas Lieutenant Governor Dan Patrick directed State senators to explore tying funding to performance on standardized tests. Texas Senator Paul Bettencourt, a member of the Senate Education Committee, was quoted by the Texas Tribune in response to the Lt. Governor's advocacy of higher stakes: "I think that in the 21st century we should be looking at other markers of success besides just showing up" (Bettencourt, quoted by Collier in The Texas Tribune, 2016).[32] High stakes policies, such as the ones discussed by Texas senators in 2016, produce compelling reasons for schools, and their districts, to create punitive measures in order to incentivize teachers to improve student performance on statewide standardized assessments. Beyond the previously discussed concerns that high stakes tests push teachers to align instruction with the content of the assessment, Berliner and Nichols raise an issue meriting additional attention. The stakes tied to performance on assessments, when those assessments serve as the single, heaviest indicator of performance, they found, stimulate practices to evade the punitive consequences of unsatisfactory student scores. Further, Berliner and Nichols found that these practices corrupt the data on student performance (Berliner and Nichols, 2005). They explain that, "in high-stakes testing environments, the greater the pressure to do well on the tests the more likely is the meaning of the score obtained by students or schools uninterpretable" (Berliner and Nichols 2005, 5).

[32] K. Collier, "Lawmakers Look at Tying School Funding to Performance." The Texas Tribune. Aug. 3, 2016. (See: https://www.texastribune.org/2016/08/03/senators-examining-performance-based-funding-schoo/)

Berliner and Nichols point, specifically, to cheating (both on the sides of the students and, perhaps worse, the teachers and administrators) and misrepresentation of important data, such as test scores and student dropout rates.[33] To illustrate the corruptibility inherent to the current environment of high stakes testing and its alignment with Campbell's law, they have relied on a compilation and analysis of news reports throughout the nation that point to instances of corrupted student performance data.

The analysis revealed that instances of teachers, principals, and districts permitting student cheating, cheating on behalf of the students, or evading accountability by finding ways to prevent the scoring of lower-performing students' exams are pervasive across the nation.[34] The issue is no longer an expected deviation from the norm: it is becoming pervasively normal. Berliner and Nichols outline eighty-three instances of alleged cheating from within over thirty different states in the US – and these are only instances that have been discovered and reported in the news. The reports include, among others, cases of teachers providing unauthorized assistance during benchmark and graduation requirement tests, teachers exposing and teaching exact test questions before the administration of exams with those questions, administrators changing student ID numbers to prevent the state recording those students' scores, and higher level boards pressuring teachers to cheat on state reported exams.[35] When schools avoid or

[33] Berliner and Nichols list in totality the corrupting effects of high stakes testing as follows: administrator and teacher cheating, student cheating, exclusion of low-performance students from testing, misrepresentation of student dropouts, teaching to the test, narrowing the curriculum, conflicting accountability ratings, questions about the meaning of proficiency, declining teacher morale, and score reporting errors (Berliner and Nichols, ii-iii).

[34] See Chapter III, Section i, subheading "*Implications for Racial and Income Equality*," for discussion on these practices.

[35] See: "Table 1: Instances and Allegations Cheating by School Personnel" (Berliner and Nichols 2005, 27-51)

distort results of the student performance to which stakes tied, 'accountability' loses validity as an impetus for punitive consequences of performance on statewide standardized assessments.

Berliner and Nichols clarify that their intent in illuminating the corruptibility of the high stakes system that manifests in teacher and district "cheating" is not to point the finger of blame at those in the teaching profession; rather they find the culpability to rest with those who have imposed a system that implicitly encourages – and creates high pressure for – such practices. They ask the important question: "why have our politicians and lawmakers created a system that pressures people who we expect to be moral leaders of our youth?" More specifically, "what sort of education system would back teachers and administrators into such a tight corner that they would cheat to ensure they have work in the future?" (Berliner and Nichols 2005, 23). Berliner and Nichols suggest that, instead of acting in service of their own interests (such as merit pay or job security), the dishonest reporting practices of teachers could – perhaps – be likened to the actions of individuals that serving a moral, albeit illegal, purpose: to act in discordance with laws that deny equality to groups of individuals.[36] For example, in North Carolina a principle refused to adhere to requirements to test all children in the school; she chooses not to test, specifically, special education students, because she sees only the potential for harm (of the special educated students) by "humiliating" the students with inevitable failure of a test already beyond their abilities (Berliner and Nichols 2005, 24).

The study reveals a necessary element of future accountability systems: in addition to lowering stakes on standardized assessments, multiple indicators of student performance must be

[36] " . . . It became plausible to us that teachers and administrators are acting no different than those who have not reported all their income to the IRS, allow prayers to be said in the schools, or defy laws that deny full equality for ethnic and racial minorities or woman. In each of these disparate cases sizable numbers of people, some liberal and some conservative, decide that their government is wrong and find justification to break laws they consider unfair" (Berliner and Nichols 2005, 24).

used to hold districts, school, and teachers accountable for providing quality education to all students. To address this finding, Federal PR 2 proposes a federally funded grant awarded to states to incentivize "intelligent accountability," including statewide development of multiple indicator systems of accountability. Additionally, State PR 1 proposes using random sampling technology in order to maintain statewide standardized tests as an indicator for performance without the high stakes, and the Local PR proposes a performance-based form of assessment that would similarly constitute an additional indicator of performance in accountability systems.

PART II

The preceding portion of this paper (Part I), dealt with issues, consequences, and subsequent principles emerging from study of past and present systems of accountability in the United States. Those principles (EPPs) demand attention from lawmakers as future legislation makes reformative changes to existing education policies and accountability systems. As the five EPPs have demonstrated, equity and quality of education depend on addressing the problems evident within current policy.

The following portion of this paper (Part II) proposes policies at the federal, state, and local levels of government to address the issues discussed in Part I. While individual recommendations are proposed within specific levels of government, these policy recommendations (PRs) are intersectional. Each proposed recommendation, in some way, relies on a separate PR, often at a different level of government, in order to obtain maximum success upon implementation. Each of the following chapters will open with an overview of the role played by that level of government in education policy, at points touching on the intersecting areas of that level with other levels of authority. The complex structure of U.S. government and its laws regarding education – necessary protections at the federal level, decentralized authority at the state level, and district oversight at the local level – require that future reforms interact cohesively and cooperatively at the intersection of varying levels of authority, in order for future policies to benefit all students.

IV. Policy Recommendations at the Federal Level

This chapter presents and defends two recommendations for education policy reform at the federal level, each of which builds on the EPPs discussed in the previous chapters. The first reform (Federal PR 1) recommends reauthorization of the Elementary and Secondary Education Act (ESEA) with lower stakes than those set in the guidelines of its current reauthorization under the Every Student Succeeds Act (ESSA). This proposal addresses EPPs 3 and 4, on high stakes and the corruptibility of data in accountability systems with single indicators of student performance, respectively. The second proposed reform (Federal PR 2) recommends eliminating the Race To The Top (RTTT) grant, and replacing RTTT with a grant that awards states for developing and implementing systems of "intelligent" accountability. Federal PR 2 addresses EPPs 2, 4, and 5, by incentivizing statewide development of systems that facilitate higher-order learning, lower stakes, and employ multiple indicators (order respective to the three listed EPPs). Because recommendations at the federal level have a specified governing body – the United States government – these recommendations are largely *critical* in nature. Federal PRs 1 and 2 assess current federal policy and conclude, informed both by evaluation of the *specific* existing legislation and EPPs, recommendations for necessary reforms. The following section discusses the federal role in education policy, to provide a frame within which to consider implementation and scope of impact for education policy recommendations at the federal level.

i. The Federal Role in Education Policy

Constitutionally, the federal government does not have the authority to legislate education on a national level. Education policy has thus, for most of the country's history, been a

state issue. The federal government has consequently remained largely uninvolved in education at the state level until the later portion of the twentieth century (Ravitch and Loveless 2000). However, with the passage of the Elementary and Secondary Education Act (ESEA) in 1965 under the Johnson administration, the federal government gained a substantial position in education policy. The creation of Title I funding established federal power of purse extending over education at the State and local levels. The legislation was a long time coming: two decades earlier in 1946, the U.S. Senate considered a bill that would create a system of federal aid to address the apparently declining[37] state of US education. This decline was especially apparent compared to other Organization for Economic Cooperation and Development (OECD) countries; the U.S. ranked second in 1995, but by 2010, it had fallen in ranking to thirteenth out of twenty-five countries (Jennings 2015, 8). Simultaneously, perception of US education became increasingly aware of systemic inequities, developing a connection between issues of education and issues of civil rights (Jennings 2015). This drew attention to the necessity of aid at the federal level. The emerging issue at this point was the gap in achievement between White and minority students. Conservatives on the right, meanwhile, blocked attempts to set federal legislation, favoring in instead the status quo – traditionally, education is purely a State issue.

It was not until Robert Taft became a serious contender for the Republican presidential nomination in 1952 that policy calling for federal involvement in education policy, by the distribution of federal funding, became remotely possible. When Taft lost the Republican nomination to General Dwight D. Eisenhower, another decade would pass before federal aid in

[37] The issue, at this point, was not a broad-scale decline, but rather a declining state of education for low socio-economic level and minority students. The achievement gap was becoming increasingly apparent, though not yet blatantly obvious, in education circles. The comparatively low performance among this population of students pulled down national averages, which were previously weighted heavily with White and middle-class students, and lent toward an appearance of nationally declining performance (Jennings 2015).

education again rose to legislative priority, this time under the Kennedy administration. After Kennedy's assassination, Lyndon B. Johnson co-opted and backed Kennedy's education agenda, folding it into his War on Poverty and clearing the way for passage of the ESEA in 1965 – the first significant federal involvement in education policy for the system of education in the United States. The keystone of ESEA, and how the federal government to this point maintains some control over education policy at the State and local levels, is Title I: the federal provision distributing funds to states, districts, and schools based on the percentage of low-income students.

Evidence of the extent to which issues in education existed, both in terms of the achievement gap and the international comparison of US performance with that of other nations, emerged with the Coleman Report in 1966, a survey of student performance requested by the Civil Rights Act of 1964. The report was used to regularly support new education legislation, and this usage combined with policies set forward in ESEA to significantly narrow the achievement gap in the following two decades (Lee, 2002).[38] The state of education in the US, at that point, seemed to have at long last established a path of improvement – ESEA fulfilled its intention, and the integrated measurement of all students, including demographics previously excluded from data, forged equity in the classroom. Jaekyung Lee, Dean of the Graduate School of Education at the University of Buffalo, argues that this improvement trend reversed, however, with the new forms of education reform initiated in the late 1980s[39] – so began the "excellence

[38] "Since the Coleman Report in the 1960s brought attention to racial inequity in student outcomes, the achievement gap between White and minority students has raised a multitude of concerns and resulted in a significant body of empirical research (see Coleman et al., 1966; Jencks & Phillips, 1998; Jones, 1984; National Center for Education Statistics [NCES], 1995; Peng & Hill, 1995)." Jaekyung Lee, *Racial and Ethnic Achievement Gap Trends: Reversing the Progress Toward Equity?*

[39] "Despite the many challenges to improving racial and ethnic equity in learning outcomes, substantial success in narrowing the Black-White achievement gap has been realized since the 1960s. However, a closer examination of the data reveals that this earlier progress has been reversed since the late 1980s." Jaekyung Lee, *Racial and Ethnic Achievement Gap Trends: Reversing the Progress Toward Equity?*

era" in education. Federal PR 1 will discuss the implications these fluctuating student performance trends have for excellence era reform.

Following the Reagan-era National Commission on Excellence in Education's release of *A Nation at Risk* in 1983, a new wave of bipartisan support emerged to re-prioritize issues in education. The report gave birth to the "excellence era" in education reform: focus in policy began to shift from need-based to merit-based in a nationwide call for high standards, high stakes, and prolific accountability testing to produce achievement. Support for this agenda mounted over almost two decades, culminating in congressional reauthorization of ESEA in 2001 with the addition of the Bush administration's standards-based education reform legislation, No Child Left Behind (NCLB). It did not take long for NCLB's support to turn to widespread opposition (Loveless 2006),[40] with critics arguing that the legislation led to a system in which teachers were encouraged, and at times forced, to "teach to the test." As the data in previous sections indicates, the attacks were well founded: the federally mandated standardized testing and stakes, in the form of punitive sanctions, tied to student performance on those tests precipitated what those in many education circles called a nationwide disaster in education. What followed was an intensified tension between the responsibility at the federal level to protect disadvantaged Americans by addressing issues of inequity, and the primarily conservative push for education to return to and remain within state jurisdiction. This tension resulted in a legislative compromise drawn during the Obama administration in 2015, attempting to reconcile blaring issues created by NCLB: the current reauthorization of ESEA under the Every Student Succeeds Act (ESSA).

---

[40] "When President Bush signed the bill into law in January 2002, Senator Edward M. Kennedy stood by his side. Four years later, NCLB faces stiff resistance from state and local authorities. Ironically, given the bipartisan support for the law, the rebellion against NCLB also seems to come from both Democrats and Republicans— from the political left and the right" (Loveless, 2006).

Due to the decentralization of education in the US, states are not *constitutionally* required to comply with ESSA regulations. While the portion of funds coming from federal sources is not large,[41] all fifty states currently comply, in part to maintain federal funding. Federal authority in education, then and now, is thus limited to power of purse. As need for funding and resources at the state level grows, this power grows. Currently the federal purse leveraging power over state and local education policy primarily consists of federal funds within Title I of ESEA/ESSA and the Race to the Top (RTTT) grant. High-stakes embedded within regulations for eligibility to receive these funds combined with growing need for funding, has created a top-down authority in education that does not intuitively coalesce with the US's decentralized structure of education. This has, as has been observed with past federal reforms, resulted in a collision course of policy. While ESSA is a step in the right direction, it does not remedy the issue: As the following section will discuss, ESSA retains high-stakes that limit state and local policy. Consequently, Federal PR 1 will explore the elements of ESSA that prevent the progressive education reform necessary to facilitate a paradigm shift at the state level.

[41] In 2012-2013, only 13% came from the federal level, where as 45% came from state sources and 46% from local sources (NCES Common Core of Data, CCD).

ii. Policy Recommendation 1: Reauthorize ESEA with Lower Stakes

Figure 4: Federal Policy Recommendation 1 (Federal PR 1)

Policy Level	Recommendation	EPP's Addressed	Means for satisfying EPP
Federal	Retain ESEA and reauthorize with lower stakes attached	EPP 3	Maintains focus on setting standards for performance and addressing need via Title I program
		EPP 4	Lowers the high-stakes tied to current reauthorization of ESEA under ESSA, allowing States to develop better systems of accountability

Few lawmakers, at this point, deny existence of a crisis in US education: namely, the stark gap in achievement between white majority students and African-American and Hispanic minority students. In 2013, African-American students, for example, still performed in the 19th percentile of white performance for reading, and in the 22nd percentile for math.[42] What many lawmakers seem less inclined to admit, however, is the futility in the current direction of standards-based policy.[43] This direction in policy has mounted over several decades: Support for what would become high-stakes, standards-based reform first took root with the Reagan-era release of *A Nation at Risk* in 1983. However, the report's diagnosis – that the nation faced a crisis of "mediocrity" in education – erroneously attributed the decline in national student performance averages to worsening overall education. The report did not account for the

[42] See note 20

[43] Reform introduced under the Obama administration, for example, even in working to rectify issues encountered under NCLB, has retained (even raised) the high-stakes tied to setting and measuring accountability standards via statewide standardized tests. We see this both in ESSA and, perhaps to a larger extent, in the Race To The Top grant (both pieces of legislation to be discussed further in later parts of this paper).

integration of previously unmeasured demographics – minority and low-income students – in the national assessment of student performance. This triumph of civil rights ensuring measurement of all student demographics to demonstrate performance naturally (initially) pulled student performance averages down. Because the performance of minority and low-income demographics were then factored into overall evaluations of performance, these demographics began to see increasingly equitable treatment in the education system. Subsequently, the achievement gap began to narrow. The *Nation at Risk* report, however, espoused a comparison between the pre-ESEA high average national performance (when most, but not all, students were measured) and the 1970's to 1980's lower performance (when all students were measured), and policymakers used this to push reform that altered the education system. So, instead of continuing to work toward equitable access to funding and resources for marginalized demographics, the report spurred an agenda that focused primarily on ripping apart and restructuring the system in place in its entirety. This was the beginning of standards-based reform, and the "excellence era" in education; NCLB, for example, attempted to use high-stakes and punitive sanctions to force achievement. The marginal improvement in performance averages that followed this reform is suggestive less of achievement and more of increasing class time spent teaching students how to take the standardized tests.[44] The failure of excellence era reforms is evident in the widening of the achievement gap, indicating that standards-based reform is responsible for reversing the previously improving performance trends among minority and low-income students.

[44] In 2012, Walter Stroup, an associate professor at the University of Texas College of Education, testified to the Texas legislature that, "the scientific basis behind the widely held suspicion that what the tests measured was not what students have learned but how well students take tests." See: Stanford, Jason. "Mute the Messenger," Texas Observer. Sep. 3, 2014. (https://www.texasobserver.org/walter-stroup-standardized-testing-pearson/)

ESSA, the current reauthorization of ESEA, is often lauded by policymakers on both sides of the political spectrum as a triumph over partisanship – a usage of compromise to retain accountability while distancing from standards-based reform and the high stakes that give birth to over testing. ESSA does, in several respects, move in a positive direction from the rigid policies set under NCLB: the 2015 reauthorization returned the power to set standards to the states and expanded flexibility for creation and usage of Statewide assessments. A close analysis of the reform, however, reveals that ESSA retains many high-stakes elements that will likely continue a culture harming teachers and marginalized student demographics. In addition to harmful byproducts at the classroom level, the continuation of high-stakes and standards-based policy limits progressive education reform at the State and local levels, creating a collision at the intersection of policy: ESSA prevents the increasingly necessary paradigm shift at lower levels of policy. Thus, Federal Policy Recommendation 1 (Federal PR 1) argues to reauthorize ESEA with a set of regulations that exhibit significantly lower-stakes than those outlined in ESSA.

ESEA took action to promote equity in education, that ultimately had a positive, narrowing effect on the achievement gap. Recognizing, then, where to make policy changes requires discerning those policies that had a negative effect on student performance and the achievement gap. The emergence of standards-based reforms, as reauthorizations of ESEA such as the Improving America's Schools Act (IASA) and NCLB shifted funding from a need-basis toward a meritocracy, aligns with reversing trends in student performance, as the achievement gap began to widen again. In an evaluation of 2013 National Assessment of Educational Progress (NAEP) Math and Reading averages for black and white students, the achievement gap has only marginally narrowed since the Coleman Report on 1965 averages in student performance. The comparison indicated that among US regions the only tangible (and still minimal) narrowing of the gap occurred in the South, where in 1965 the achievement gap was

the widest. Worse still, the achievement gap in Midwest region NAEP scores remained unchanged for Math and widened slightly for Reading performance averages.[45] This indicates that, since reform began attempting to address the achievement gap in the late 1960s, with legislation allegedly targeted specifically at narrowing the gap while also facilitating overall national improvement, little to no progress has been made to equalize performance and opportunity among different demographic groups of students. Yet, while significant improvements in narrowing the achievement gap do not exist when comparing progress between 1965 and 2013 NAEP scores, it can be determined that ESEA's original need-based policies had a positive effect on the achievement gap. A shorter-term evaluation of progress from the 1965 scores at the beginning of modern education reform and scores in the late 1980s and 1990s indicates that improvement took place *before* standards-based reforms rose to the forefront of policy initiatives – but not after (Anrig 1985, Lee 2001).

In a comparison of NAEP Math and Reading averages for black and white students between the early 1970s and the late 1980s, the achievement gap *narrowed* by 20% to 40% across the US – black students exhibited considerably improved performance, while white student performance remained steady (Lee, 2001). In the mid-to-late 1980's, however, the trend of progress began to reverse, and the achievement gap returned – as demonstrated by the 2013 evaluation of the achievement gap on the NAEP – to black-white student performance deviations very close to or surpassing it's the gap's previously reported width in 1965. Lee clarifies the

[45] Education Next, *The Black-White Achievement Gap Persists (Figure 1)*: "In both math and reading, the national test-score gap in 1965 was 1.1 standard deviations, implying that the average black 12th grader placed at the 13th percentile of the score distribution for white students. In other words, 87 percent of white 12th graders scored ahead of the average black 12th grader. What does it look like 50 years later? In math, the size of the gap has fallen nationally by 0.2 standard deviations, but that still leaves the average black 12th-grade student at only the 19th percentile of the white distribution. In reading, the achievement gap has improved slightly more than in math (0.3 standard deviations), but after a half century, the average black student scores at just the 22nd percentile of the white distribution."

direction of the averages for black and white students when evaluated independently to determine the movement, aside from the width, of the achievement gap:

> "During the period between 1971 and 1986/1988 when the achievement gap between Whites and Blacks narrowed, White students' achievement level was quite flat, whereas Black students made substantial academic gains. In contrast, during the period between 1986/1988 and 1999, when the gap grew, the pattern reversed: White students improved their achievement but Black students made few gains on NAEP. Consequently, the narrowing of the Black-White achievement gap stopped, and in some cases the gaps returned to the level of the late 1970s or early 1980s. Time-series regression analyses show a significant quadratic (downward slope followed by upward slope or leveling off) trend of the Black-White achievement gap for all age groups in mathematics for ages 9 and 13 in reading. The turning point in the Black-White gap trend appeared to be in the late 1980s and early 1990s." (Lee, 2002)

The direction of trends in student performance and the achievement gap both indicate the influence that specific policies have had on those trends. As narrowing trends in the achievement gap appeared to reverse – in the late 1980s and the early 1990s – education reform in the US began the period of time known as the "excellence era."[46] During this period, focus in policy shifted from need, where emphasis remained on aid programs such as Title I, to a meritocracy, with an emphasis on standards and accountability.[47] (Lawmakers officially saw to this in 1994, when the first standards-based reform was introduced within the IASA reauthorization of ESEA.) Aligning policy with the initial narrowing and later widening trends in the achievement gap, indicates that the "excellence era" reforms were not only unsuccessful, but harmful. Further, evidence points to the high-stakes nature of the reform standards as the root of the issue (Berliner 2005).

---

[46] "Dubbed the 'excellence era' in US educational reform, the period from 1980 to the turn of the century delineates a generation of educational policies intended to enhance student learning." (Murphy and Adams, Jr., 1998; p. 426)

[47] "Early reform initiatives of this era indicate that the government reform portfolio is dominated by efforts to develop standards and accountability mechanisms." (Murphy and Adams, Jr., 1998; p. 434)

ESEA's (in its original passage) narrowing effect on the achievement gap demonstrates the beneficial value of the federal legislation: The comparison of Black-White performance averages between the 1960s and the 1980s ultimately suggests that ESEA contributed toward a more equitable system of education. Under the initial years of ESEA, achievement among black students rose significantly, while achievement among white students remained steady.[48] Intervention in education policy at the federal level, even in the decentralized system of US education, thus can and *should* play a critical role in ensuring equity for students across the nation. For this reason, steps taken at the federal level to address current issues in education should retain ESEA as the broad policy legislating the federal position in education, due to the data indicating ESEA's value and effectiveness as an element of education policy capable of producing positive trends on the national level. This policy recommendation thus focuses on the current reauthorization, ESSA, as the area necessitating policy change.

The conclusion of Federal PR 1 thus follows that the federal government should reauthorize ESEA under new legislation for accountability that lowers the stakes currently in place under ESSA. The basis for this recommendation lies in the continuation of high-stakes at the federal level as written in ESSA. EPP 3, "High Stakes are Detrimental in Systems of Accountability," demonstrates the need to reduce stakes: A new, lower-stakes reauthorization of ESSA will enable effective policy reform at the State and local levels. The following table presents an analysis of individual elements of ESSA that retain the high-stakes nature of policy, demonstrated by EPP 3 to be detrimental to students, particularly low-income and minority

[48] Critics of need-based legislation often claim that it serves to disadvantage high-achieving demographics. "Steady" performance, in this case, does not merit criticism: The White, steady-income demographic of students exhibited performance well aligned with (or surpassing) any "challenging" standards set under NCLB or other high-stakes legislation. Education accessed by this demographic appeared, then, to satisfy federal, State and local goals.

demographics. Included in addition to an analysis of the continued high-stakes nature of the regulations are recommendations regarding measures to lower stakes set at the federal level.

Figure 5: Analysis of Stakes in ESSA and Subsequent Recommendations

The following table evaluates individual regulations set forth in ESSA with regard to the nature of the stakes they create, and the continuing emphasis placed by each of the outlined regulations on statewide assessments. Those regulations that have been included in this table have been identified as high-stakes in nature. In addition to an analysis of how the regulations create or retain the stakes set at the federal level, the table includes recommendations for scaling back these stakes in a new reauthorization, as recommended in Federal PR 1. All excerpts included to describe the regulation have not been taken from the written legislation, but from the outline of regulations included on US Department of Education website.

Regulating Element within ESSA	Analysis of Stakes	Recommendation (Fed. PR 1)
"The law maintains the requirement that states administer to all students annual statewide assessments in reading/language arts and mathematics in grades 3-8 and once in high school, as well as assessments once in each grade span in science for all students and annual English language proficiency assessments in grades K-12 for all English learners." *(ESSA, Title I)*	ESSA upholds the **annual** requirement for statewide standardized testing, which contributes to a continuation of the current yearly (and yearlong) emphasis on test preparation within the classroom.	Replace the annual requirement with assessment periods over multiple years, based on assessing progress within broad developmental stages rather than grade-levels.
"ESSA requires that states establish college-and career-ready standards and maintain high expectations when assessing all students against those standards." *(ESSA, Title I, Part A)*	This retains the NCLB-era focus on "college and career ready" standards, which do not necessarily align with learning indicators.	Eliminate the requirement of a "college and career ready" framework for standards, and allow States to fully define and frame standards.

"States must assess all students." *(ESSA, Title I, Part A)*	The implications of this element of ESSA are nuanced: this requirement is positive in that it seeks to promote equity. However, assessing every student every year maximizes exposure to standardized testing – and thus the potentially harmful byproducts – for each student.	Equity can be achieved without requiring the assessment of every student every year. Every demographic must require assessment, not every student, and students within demographics must be selected for assessment for accountability purposes using statistically random methods.
"A state may permit districts to use a nationally recognized[49] high school academic assessments in place of the statewide high school assessment; a district using this flexibility, however, must use the same locally selected, nationally recognized assessment in all of its high schools." *(ESSA, Title I, Part A)*	This approaches a step in the right direction by providing an alternative to statewide assessments, but limits those alternatives: "nationally recognized" assessments include only assessments in the nature of the SAT and ACT. This provision does not provide flexibility for States or LEAs to design those alternatives, thus sustaining the emphasis on and ease of usage of statewide assessments.	Replace "nationally recognized requirement" with a peer review process to provide flexibility for individualized development of alternatives to statewide assessments; maintain requirement that States prove assessments align with state-set standards.
"To ensure that the vast majority of students take a state's general assessment and only students with the most significant cognitive disabilities take an alternate assessment aligned with alternate academic achievement standards, the ESSA limits the number of students who may take such assessments to 1 percent of all tested students in a given subject." *(ESSA, Title I, Part A)*	Places undue emphasis on statewide-standardized assessments by limiting State usage of alternative forms of assessment, both in terms of demographics and number of students allowed to take alternate assessments.	Substantially raise or eliminate the 1 percent cap on students allowed to take alternate assessments, and open alternate assessments to all students, with the requirement that equal number of students in each demographic be tested using alternatives.
"The Department may grant innovative assessment demonstration authority to up to seven states during the initial demonstration period of three years." *(ESSA, Title I, Part B)*	Places undue emphasis on statewide-standardized assessments by limiting State usage of innovative, alternative forms of assessment.	Eliminate or raise the seven state limit on the number of states granted innovative assessment demonstration authority.

[49] "To ensure these tests are truly 'nationally recognized,' the regulations clarify they must be given in multiple states, be recognized by institutions of higher education for the purposes of entrance or placement into courses in postsecondary education or training programs, and provide the same benefits to all students – including English learners and children with disabilities."

"In order to provide ample time for the transition to new statewide accountability systems, particularly to allow for meaningful stakeholder engagement and thoughtful inclusion of new accountability indicators, the final regulations give states until the 2018-19 school year to identify schools for comprehensive and additional targeted support and improvement, with annual identification of schools with consistently underperforming subgroups for targeted support and improvement beginning in 2019-20." *(ESSA, Statewide Accountability Systems)*	With ESSA signed in 2015-16, the 2018-19 deadline for identification of comprehensive and targeted improvement schools has imposed a still narrow turnaround timeline, considering the new nature of the regulations, for schools hoping to avoid intervention (3-4 years). The stakes of this element likely produced heavy emphasis on indicators for improvement in schools, which at this point largely remain statewide-standardized assessments. The annual identification following this perpetuates this practice, perhaps to a larger degree due to the annual requirement.	Lengthen annual identification for improvement intervention to biennial identification or identification every three years. This also creates flexibility for statewide assessment indicators include in this factor of accountability, such as standardized testing, to be given in developmental periods rather than annually.
"The final regulations require that each state's accountability system meaningfully differentiates schools by providing them with a summative determination from among at least three distinct, clear, and understandable categories, while allowing for multiple ways for states to designate schools in each category." *(ESSA, Statewide Accountability Systems)*	The summative determination requirement retains the meritocratic measurement of schools, rather than the differentiation of schools based on need, by measuring relative density of Title I students.	Shift differentiation mechanisms to rely primarily on a need-basis, utilizing percentage of Title I students, rather than summative determinations that enable meritocratic distribution of funding and resources or improvement intervention determinations.
"States must demonstrate that schools that would have been identified for comprehensive support and improvement on the basis of 'substantial' indicators, but are not identified on the basis of the state-selected indicators taken together, have made significant progress for all of their students on at least one 'substantial' indicator." *(ESSA, Statewide Accountability Systems)*	Despite appearing to return authority to States to select and weigh indicators, this legislation ties stakes to determination of 'substantial' indicators. This retains federal control over indicators.	Return federal control over determination of weight for indicators to State and Local Education Agencies.
"States must demonstrate that schools that would have been identified for targeted support and improvement on the basis of 'substantial' indicators alone, but are not identified on the basis of all state-selected indicators taken together, have made significant progress for the subgroup that is struggling on at least	Despite appearing to return authority to States to select and weigh indicators, this legislation ties stakes to determination of 'substantial' indicators. This retains federal control over indicators.	Return federal control over determination of weight for indicators to State and Local Education Agencies.

one 'substantial' indicator." *(ESSA, Statewide Accountability Systems)*		
"The law requires that all students take statewide assessments and that states factor into their accountability systems participation rates below 95 percent for all students or subgroups of students." *(ESSA, Statewide Accountability Systems)*	The 95 percent participation requirement for statewide assessments retains the strong federal emphasis on State mandated usage of statewide-standardized assessments. This also threatens States with growing opt-out movements among parents and stakeholders (such as New York[50]) from losing their federal funding for falling below 95 percent participation.	Significantly lower the participation percentage cap, or allow the participation percentage to account for participation on non-statewide systems of assessment in addition to statewide assessments.
"Under the statute and the final regulations, states must identify certain schools at least once every three years for *comprehensive* support and improvement, including: at least the lowest-performing 5 percent of Title I schools in the state; high schools with graduation rates at or below 67 percent (or a higher percentage selected by the state) for all students based on the four-year adjusted cohort graduation rate; and Title I schools with chronically low-performing subgroups that have not improved after implementing a targeted support plan for a state-determined number of years. *(ESSA, Supporting Low-performing Schools)*	Requiring identification for large-scale improvement with short three-year turnaround deadlines ties high-stakes to the listed indicators, including performance and graduation rates.	Extend the improvement turnaround deadline and return emphasis to need rather than meritocracy.
"States must also identify schools for *targeted* support and improvement, including: schools with a subgroup performing similarly to all students in the lowest-performing 5 percent of Title I schools, to be identified each time the state identifies its schools for comprehensive support (these schools must also receive additional targeted support); and schools with a consistently underperforming subgroup, as defined by the state, annually." *(ESSA, Supporting Low-performing Schools)*	Requiring identification for large-scale improvement with short three-year turnaround deadlines ties high-stakes to the listed indicators, including differentiation of subgroups, which can promote emphasis on statewide assessments among those subgroups rather than enabling equity across all classrooms.	Extend the improvement turnaround deadline and return emphasis to need rather than meritocracy.
"[The regulations] will also ensure that states and districts set meaningful exit	Exit criteria and threat of additional action ensures high-stakes tied to	Extend turnaround deadline to meet exit criteria and

[50] See: Harris, Elizabeth A. "20% of New York State Students Opted Out of Standardized Tests This Year" *The New York Times*. Aug. 12, 2015. (https://www.nytimes.com/2015/08/13/nyregion/new-york-state-students-standardized-tests.html)

criteria for identified schools, including requiring additional action where initial interventions do not work to improve student outcomes." *(ESSA, Supporting Low-performing Schools)*	performance within identified schools, which combines with short improvement-turnaround deadlines to retain emphasis on performance indicators such as statewide assessments.	ensure that "meaningful exit criteria" includes several indicators beyond performance on statewide assessments to reduce focus on standardized test preparation in identified schools.

The above breakdown of ESSA guidelines indicates the extent to which the reauthorization retains a significantly heavy emphasis on standards-based policy and intervention, due to the apparent high-stakes nature of many included regulations. Thus, ESSA continues the same policy enabling a culture of fear that falls from the top down to States, LEAs, schools, teachers, and ultimately students. These effects, just as with NCLB, will have particular impact on schools identified as low-performing, which due often to zoning and lack of resources and funding are composed primarily of minority, low-income students. Students from these demographic profiles will consequently shoulder the bulk of harmful byproducts of high-stakes: the same overexposure to statewide-standardized assessments that limit learning practices extending beyond rote mechanisms.

Regrettably, what should be the most promising action of partisan compromise included in ESSA, intended to correctly reduce emphasis on over testing, has ironically enabled what is perhaps the legislation's most dangerous effect: By returning some federal control in specific determination of standards to States and LEAs while simultaneously retaining the high-stakes tied to performance alignment with those standards, ESSA unintentionally incentivizes action to *set lower standards*. The optical nature of ESSA as a political compromise has ironically resulted in many of the legislation's potentially worst consequences. Conor Williams, a senior education-policy researcher for New America, articulates these concerns in identifying ESSA as a political device rather than a policy solution:

> "As far as I can tell, it's a brilliant piece of political posturing ... that doesn't seem likely to provide educational opportunity for underserved kids… It's a clear system that serves the political needs of most members of Congress and protects a variety of special interest groups. It combines a thin veneer of civil rights equity with excruciating complexity and uncertain accountability. It takes a relatively simple federal accountability system, removes the teeth, and layers on a bunch of vague responsibilities for states … Just because something is a compromise doesn't mean that it will do good things for children."

This counter-productive effect of (seemingly) positive "compromise" at the federal level demonstrates that coordination at the complicated intersection of federal, State and local policy is of critical importance as lawmakers move forward in addressing issues in education.

iii. Recommendation 2: Replace RTTT with Grant Funding to Incentivize "Intelligent Accountability" Innovations

Figure 6: Federal Policy Recommendation 2 (Federal PR 2)

Policy Level	Recommendation	EPP's Addressed	Means for satisfying EPP
Federal	Replace RTTT grant with grant rewarding intelligent accountability systems	EPP 2	Incentivizes statewide development of an accountability system that minimizes testing of lower-order skills, and maximizes assessments enabling higher-order learning.
		EPP 4	Incentivizes statewide development of accountability systems with lower stakes.
		EPP 5	Incentivizes statewide development of accountability systems with multiple indicators.

Aid for elementary and secondary education grew by approximately $100 billion in 2009 under the American Recovery and Reinvestment Act (ARRA), a measure introduced during the

Obama administration in order to stimulate the national economy during the recession.[51] In addition to setting high-stakes inducing requirements, such as stricter teacher evaluations and improvement of low-performing schools, to qualify for receipt of aid, ARRA strengthened the federal purse with the $4.35 billion competitive grant program, Race To The Top (RTTT). RTTT was designed to, "reward States that have demonstrated success in raising student achievement and have the best plans to accelerate their reforms in the future."[52] The program has persisted with yearly appropriations of funds despite the stabilization of the economy post-recession, and is the most recent, significant demonstration of the federal shift from need-based aid to merit-based aid, as policy distances from Title I programs in favor of those like RTTT which provide federal funding in response to demonstrations of successful improvement.[53]

RTTT, however, creates additional stakes tied to student performance on standardized statewide assessments, and encourages punitive measures against teachers and principals on the basis of performance on these assessments – thus contributing to a culture of fear and over testing. Evidence for this lies in an analysis of the point breakdown for determining states awarded the RTTT grant, and the value system this indicates at the federal level. Awarding of RTTT is determined on a point system divided into six selection criteria: Great Teachers and Leaders (138 points); State Success Factors (125 points); Standards and Assessments (70 points);

---

[51] "On February 17, 2009, President Obama signed into law the American Recovery and Reinvestment Act of 2009 (ARRA), historic legislation designed to stimulate the economy, support job creation, and invest in critical sectors, including education." *Race To The Top Executive Summary,* US Department of Education, November 2009

[52] The Department of Education lists four "core education reform areas" in which implementation and success would be evaluated to determine awarding of RTTT funding:

> *"Adopting standards and assessments that prepare students to succeed in college and the workplace and to compete in the global economy; Building data systems that measure student growth and success, and inform teachers and principals about how they can improve instruction; Recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and Turning around our lowest-achieving schools."*

[53] "The desire for rapid change and improvement may explain Obama's support for new reforms like RTTT and lack of interest in the traditional federal programs such as Title I. … Title I has receded into the background, and hope for school improvement is placed elsewhere" (Jennings, 53).

General Selection Criteria (55 points); Turning Around the Lowest-Achieving Schools (50 points); and Data Systems to Support Instruction (47 points).[54] Decoding the positive rhetoric of the selection criteria necessitates a closer look at the subcategories of each criterion, and the respective individual point breakdown of those subcategories. For example, the second subcategory under Great Teachers and Leaders is "Improving teacher and principal effectiveness based on performance," which is valued at 58 points, giving this category more weight in evaluation of States' applications for the grant than the whole criteria categories for General Selection Criteria, Turning Around Lowest-Achieving Schools, and Data Systems to Support Instruction. Further, the highest-weighted element (28 points) of this subcategory includes annual "rigorous" evaluations of teachers, and using these evaluations to "inform decisions regarding … removing ineffective tenured and untenured teachers and principals after they have had ample opportunities to improve, and ensuring that such decisions are made using rigorous standards and streamlined, transparent, and fair procedures." *See the breakdown in the figures on the following pages.*

[54] The selection criteria here are listed in order of descending point value, but they are listed in a differed order within the Department of Education Executive Summary of the program: State Success Factors; Standards and Assessments; Data Systems to Support Instruction; Great Teachers and Leaders; Turning Around the Lowest-Achieving Schools and General Selection Criteria.

Figure 7: Weighted Values in Race To The Top Grant

The chart below demonstrates the relative weighted values of the individual categories that constitute the point system for determination of States awarded RTTT grant funding. The maximum points allocated for each category determine the relative percentage weight of each category, as visually demonstrated below.

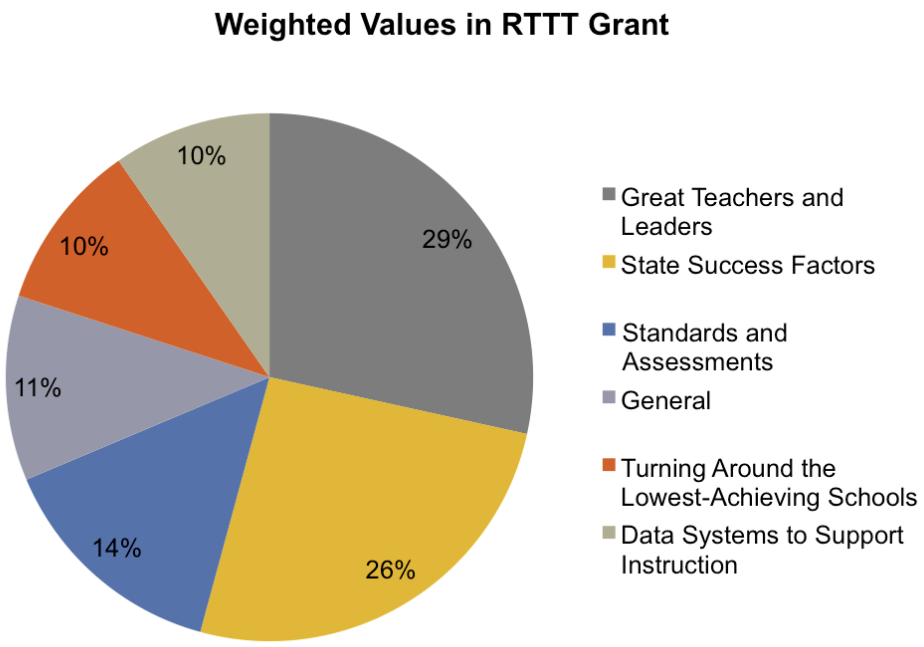


Figure 8: Weighing Subcategories of "Great Teachers and Leaders" Category in RTTT

The chart below demonstrates the relative weighted values of the individual subcategories that constitute the broader "Great Teachers and Leaders" category of RTTT. Note the substantial weight (42%) placed on the subcategory titled "Improving teacher and principal effectiveness based on performance" relative to other subcategories. The maximum points allocated for each category determine the relative percentage weight of each category, as visually demonstrated below.

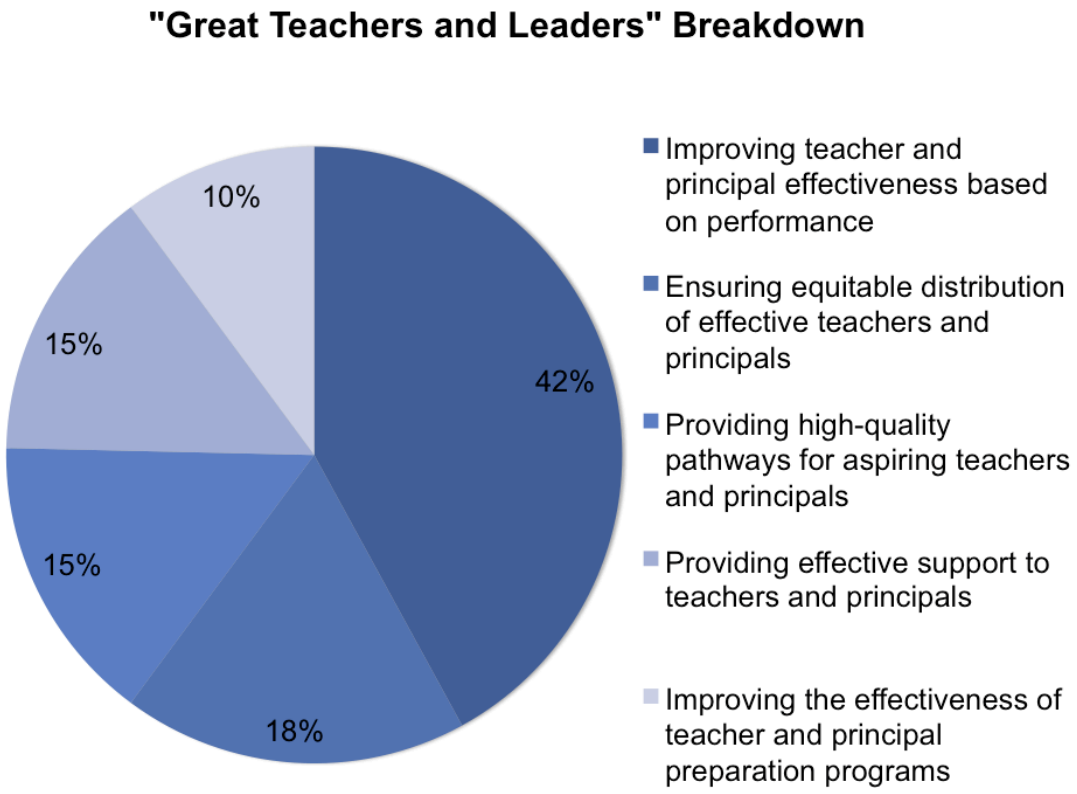
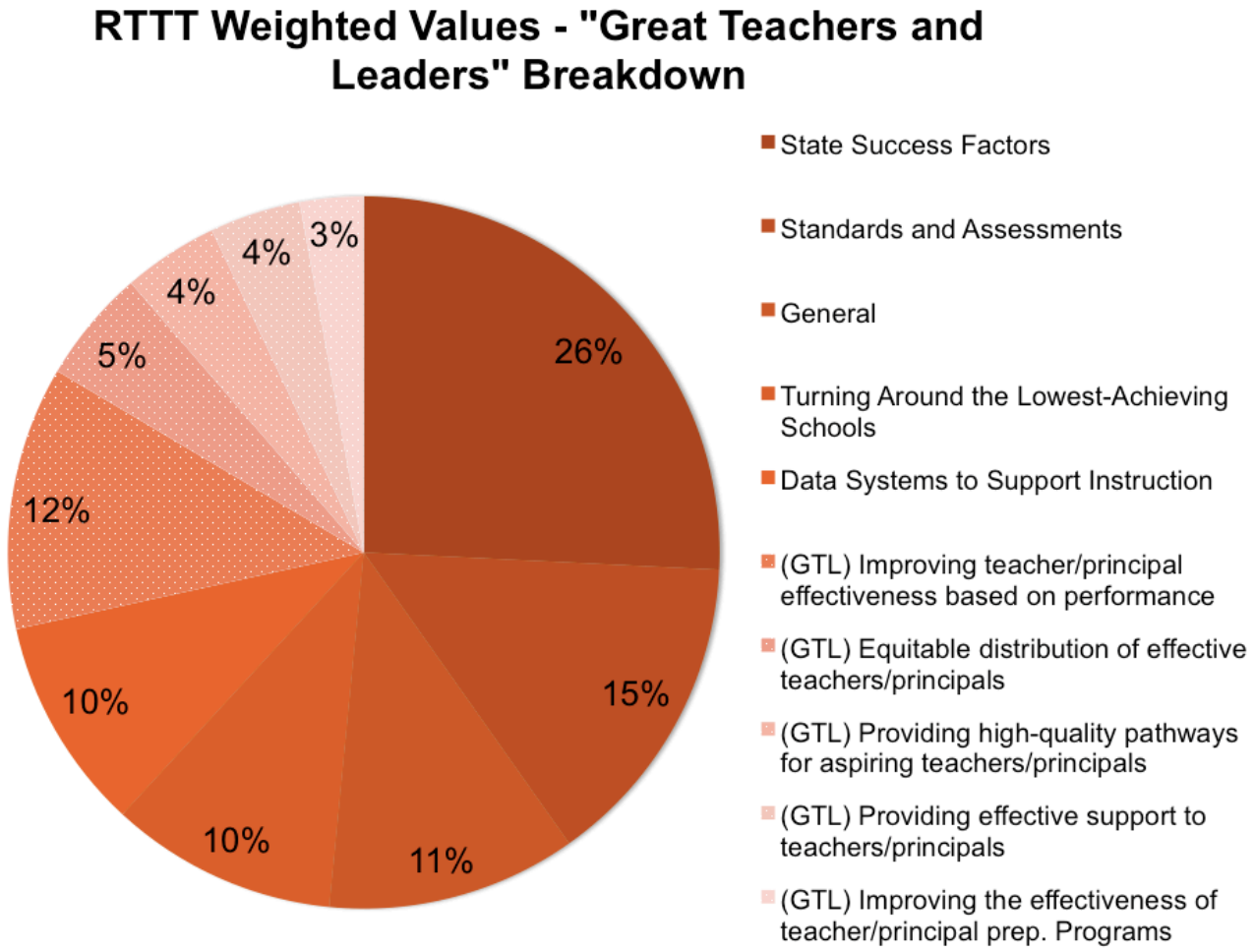


Figure 9: Weighted Values in Race To The Top Grant

The chart below demonstrates the same RTTT point breakdown of relative weighted values for broad categories as in Figure 7, but delineates one specific broad category, Great Teachers and Leaders (GTL), into its individual subcategories – GTL subcategories are differentiated by the dot pattern. This demonstrates the weight of the GTL subcategories relative to the broad categories of the RTTT point system. Note the substantial weight placed on the "Improving teacher and principal effectiveness based on performance" subcategory (12%) relative to other RTTT broad categories. This subcategory is allocated more weight than three of the six broad categories: General (11%), Turning Around the Lowest Achieving Schools (10%), and Data Systems to Support Instruction (10%). The maximum points allocated for each category determine the relative percentage weight of each category, as visually demonstrated below.



This delineation of the weighted values in the point-based evaluation process for the RTTT grant demonstrates the value system and subsequent stakes at the federal level: Inordinate weight, and thus high-stakes, is tied to a measure that evaluates and subsequently removes teachers and principals based on effectiveness in improving student performance. It is important to note that this measure is placed within a subcategory, despite its weighted value being higher than those of *half* of all individual RTTT point categories. Furthermore, it becomes necessary to recognize what, specifically these stakes are then tied to, as written in this subcategory: "student performance." Up to this point and under ESSA, "student performance" is most efficiently determined at the federal and State levels by performance on statewide-standardized assessments – currently, the focus for these standards is centered (especially by measures such as RTTT) on the Common Core State Standards Initiative (Common Core). This weight breakdown of RTTT thus creates an additional layer of stakes tied specifically to statewide assessments aligning to Common Core standards.

This weighting becomes increasingly problematic when considering both the corruptibility of high-stakes standardized assessments, as established by the research of Berliner and Nichols (2005), and the inability of these assessments to indicate true student and teacher performance, as explained by Stroup's testimony of his research. RTTT, then, places undue emphasis on evaluating teachers and principals according to student performance on standardized assessments that do not accurately indicate teacher effectiveness. Just as punitive sanctions under NCLB induced State-level action mandated strict evaluation of teachers according to student performance on test scores,[55] these weighted categories for determining grant funding continue to corrupt data by incentivizing inaccurate data reporting and stimulate emphasis on teaching test material in the classroom in order to meet standards for evaluation:

[55] See discussion in Section I

The additional stakes created by RTTT valuations is not limited to this category, or subcategory therein, and the federal grant has received its fair share of criticism:

> "Race to the Top, in a very short period of time, has set the stage to expand significantly an approach to education reform that is focused on a questionable test-based accountability model, fostering notions of choice and competition that have yet to produce the promised dramatic improvements in educational outcomes." *Urban Public Education Reform: Governance, Accountability, Outsourcing,* Natalie Gomez-Velez

RTTT ultimately constitutes federal action that bolsters the high-stakes and rigid system of accountability incorporated within and enforced by first NCLB and now ESSA, while attempting to define and nationalize standards through the Common Core. As a result, the program serves to reproduce many of the unintended effects stemming from past, highly criticized reform initiatives. So the question, instead, becomes: how does the federal government create a system of accountability that (a) produces valuable, usable data, and (b) does not place undue burden on students and teachers? These effects precipitate the following policy recommendation: The federal government should eliminate the RTTT grant, and redirect annual appropriations to an alternative competitive grant, based on a system of evaluation that rewards and encourages States to adopt and implement "intelligent accountability," or innovative methods of assessment of student performance and achievement.

The measures of evaluation for this new grant do not immediately seem of significant difference to the current aim for RTTT, which is to "encourage and reward States that are creating the conditions for education innovation and reform." The difference, then, rests in the "conditions" for "innovation" – rather than incentivizing high-stakes accountability and adoption of the Common Core standards, the weighted categories for this competitive grant should focus on metrics that reward innovations and reform initiatives successfully (a) minimizing individual student exposure to statewide standardized testing, or (b) facilitating LEA adoption of

performance-based assessments that focus on analytical skills and project-based tasks. These assessments, though somewhat subjectively measured depending on instructor interpretation of standardized rubrics, must demonstrate success in objectively measurable areas such as graduation rates and post-secondary educational attainment and completion, in order to provide evidence of efficacy.

"Intelligent Accountability" is a concept that, though encouraged by and rewarded to states via funding at the federal level, should remain flexible for states to assume nuanced implementation methods that best serve independent state and local constituencies. Accountability without high-stakes and standardization indeed constitutes a much needed paradigm shift. In a decentralized education system, this paradigm shift must occur at the state level. Consequently, State PR 1 and State PR 2 present and defend specific recommendations for this *new* accountability.

V. Recommendations for Policy Reform at the State Level

The previous chapter made recommendations for education policy that were specified toward existing legislation. Consequently, the federal PRs included analysis of direct elements of that legislation at the federal level. However, this chapter on state policy recommendations and the following chapter regarding local policy do not center on a specific local region; rather, this recommendation is proposed as a policy "ideal," for implementation within a federal jurisdiction that provides the according flexibility for this state recommendation. Thus, this PR serves as a general model for two policy reforms at the state level: the previous recommendations at the federal level aim to lower the stakes tied to accountability, which subsequently facilitates the recommendations made in this chapter at the state level. State PR 1, which recommends statewide usage of stratified random sampling techniques to measure student performance on standardized assessments, addresses EPPs 3, 4 and 5, by maintaining standards and a means for measurement, lowering the stakes associated with assessment, and constituting a multiple indicator in a system of accountability. State PR 2 recommends incorporating demographics on the basis of need into the formula to determine state funding. This PR addresses the issue diagnosed in EPP 1 by preventing a number of unintended consequences of a system favoring merit over need. Together, State PRs 1 and 2 shift paradigms for accountability toward a need-based system with lower stakes accountability.

i. Education Policy at the State Level

Because education in the United States is decentralized, a sizeable portion of power is concentrated within the State governments, particularly in terms of education policy. In turn, that authority is delegated within the State Education Agencies (SEAs) of each state. The federal

government, as discussed in the previous chapter, sets guidelines for accountability at the state level; Local Education Agencies (LEAs) implement State policies and practice accountability measures; and, somewhere in the middle, SEAs establish statewide education policies, such as accountability systems. As a result of this decentralized structure, state laws dictating education policy differ from state to state, according to their individual constitutions. The US Department of Education lists among the duties authorized by states: oversight, public education funding, licensing and certification, policies for curricula and standards, structuring the governance of state boards of education, and provision of education services for special needs citizens.[56]

Chapter IV examines the capacity of federal government to control education policy, both in the extent provided under Elementary and Secondary Education Act (ESEA)[57] and in its authority to set federal protections[58] for marginalized students. The previously extensive role of the federal government under NCLB has since lessened with the passage of the Every Student Succeeds Act (ESSA),[59] which returned a great degree of authority to the state level, by allowing states to set standards for accountability. ESSA, however, still sets guidelines for state systems of accountability. Further, as found in the analysis of ESSA in Federal PR 1, ESSA upholds many regulations that produce high-stakes accountability at the state level. Subsequently, Federal PR 1 and PR 2 included recommendations to lower those stakes. Lower stakes for accountability – via fewer or more flexible regulations, such as eliminating the 95% cap on student participation on

[56] State governments, "perform the political, administrative, and fiscal functions that are often the work of ministries of education in countries with centralized education systems. Education is the largest budget item for each of the 50 state and 5 territorial and commonwealth governments within the United States. The degree to which states and territories control education depends upon their constitutions, statutes, and regulations." See: "Organization of U.S. Education: State Role I: Primary and Secondary Education" *International Affairs Office, U.S. Department of Education*. (http://www.ed.gov/international/usnei/edlite-index.html)

[57] See Chapter V, Section i: "The Federal Role in Education Policy"

[58] These include federal laws such as the Individuals with Disabilities Act (IDEA), which protects against discrimination on the basis of disability status, and Title IX, which protects against discrimination on the basis of sex.

[59] 2015 reauthorization of ESEA

statewide assessments[60] – at the federal level consequently expands powers at the state level to implement *new* systems of accountability. Thus, recommendations at the federal level facilitate an increasingly necessary paradigm shift at the state level for accountability and stakes.

State PR 1 and PR 2 dictate the means for achieving that paradigm shift: a system of accountability that facilitates lower stakes and addresses need before rewarding merit. State PR 1 discusses utilization of stratified random sampling methods employed by NAEP assessment procedures, at the state level in statewide accountability systems. Because EPP 2 demonstrated the consequences of merit-based systems of funding, State PR 2 prescribes a need-based system of funding for the significant portion of monetary resources originating at the state level. The following sections demonstrate how these new methods for accountability will enable reliable data for performance measurement and equity in the classroom.

[60] See Figure 5

ii. State PR 1: Utilize Stratified Random Sampling in Statewide Standardized Assessments

Figure 10: State Policy Recommendation 1 (State PR 1)

Policy Level	Recommendation	EPP's Addressed	Means for satisfying EPP
State	Utilize stratified random sampling techniques for statewide standardized assessments	EPP 3	Continues standardized system for development of informative data at state level
		EPP 4	Fewer students per classroom tested each year lowers stakes tied to classroom performance on standardized assessment
		EPP 5	Provides a method for retaining statewide standardized assessments without the harmful iatrogenic effects, constituting an additional indicator in a multiple indicator accountability system

ESSA, as discussed in the previous chapter, requires State accountability systems to measure student performance with annually delivered sets of statewide standardized assessments,[61] and analysis of the specific regulations within the legislation indicate that high stakes remain tied to student outcomes.[62] Further, findings in Chapter 1 suggested that these standardized assessments, especially high-stakes assessments, often generate harmful

[61] "Many assessment provisions remain unchanged, and States must continue annual statewide tests in reading/language arts and mathematics to all students grades 3-8 and once in high school as well as in science at least once in each of grades 3-5, 6-9, and 10-12." Source: U.S. Department of Education
[62] See Figure 5, "Analysis of Stakes within ESSA Regulations"

circumstances for students, particularly those belonging to marginalized demographics. Independent research has consistently supported findings that these assessments: undermine individualized instruction in the classroom,[63] are not accurate indicators of academic achievement,[64] (Sacks 1997, 26-28) and tend to correlate more strongly with socioeconomic level than other academic indicators (Natriello and Pallas 1999; Nichols, Glass, and Berliner 2005; as cited in Moses and Nanna 2007). These findings are more problematic when combined with additional evidence that, "In contrast to literacy policy's aims to help promote the 'well-being' of all learners and 'equity' within the educational system, youth attest to feeling 'shame' and show further marginalization due to this testing mechanism" (Kearns, 112). Together, these conclusions suggest that this effect compounds for low-income and minority students. These students are shown to disproportionately perform lower in correlation with socioeconomic class – this lower performance on assessments may likely amplify the impacts of their marginalized status, creating a cyclical trend of low achievement. Attempts at equity, then, should benefit from minimizing these effects by reducing exposure to the assessments themselves. Yet simultaneously, equity in education necessitates equal standards for achievement, and measuring those standards across demographics to regulate this necessitates maintenance of performance data.[65]

Federal law dictates that the NAEP, a nationwide standardized assessment widely and reliably used to draw conclusions about student achievement across states and demographics, utilize a random sampling process to produce data representative of student performance on a

[63] "At the K-12 level especially, teachers testify that standardized tests don't accurately measure their students' abilities and that widespread practices of "teaching to the test" render test scores virtually meaningless. In 1994, Educational Policy published a study on teachers' views of standardized tests. Just 3 percent of teachers in one sample agreed that such tests are generally good, "whereas 77 percent felt that tests are bad and not worth the time and money spent on them." According to the study, about eight in 10 teachers believe their colleagues teach to the tests." (Sacks 1997, 28)

[64] Sacks, 26-27; See note 47 on Walter Stroup testimony

national and regional basis.[66] The National Center for Education Statistics (NCES) reports that the scientific sampling process employed nationally by the NAEP enables the exam to: "Make valid statements about the performance of student groups across the nation (12 million students total), even though only one–tenth of these students are assessed; Contain costs and concentrate data collection efforts; [and] Minimize the burden to the nation's school systems by only assessing a representative sample of their students" (Kolstad, *Basic Sampling Concepts Used in NAEP*). Because NAEP data is not reported for individual students, results for every student in every school need not be reported. Further, no single student selected for NAEP assessment takes the "several hundred assessment questions" necessary for "valid and reliable assessment of NAEP content."[67] The random sampling technique thus, in addition to cutting expenditures on assessments, minimizes the number of students exposed to standardized assessments in the classroom while simultaneously providing data necessary to measure achievement *with validity*. NAEP uses stratified random sampling techniques to successfully obtain this data, and decisions at various levels of policy are subsequently informed by this data – So, in the face of overwhelming evidence indicating the adverse consequences of standardized assessments on students, why has the technique not been employed at the State level? State PR 1 thus prescribes that state accountability systems use stratified random sampling[68] to produce data representative on a statewide and district-wide, or locally regional, basis.

[66] Per *NAEP Authorization Act, SEC 303 (b) (2) (A)*: "The Commissioner for Education Statistics… shall…use a random sampling process which is consistent with relevant, widely accepted professional assessment standards and that produces data that are representative on a national and regional basis." (Kolstad, *Basic Sampling Concepts Used in NAEP*, National Center for Education Statistics)
[67] EdSource.org, "Frequently Asked Questions About NAEP Sampling: Developed in support of reporting 2005 NAEP reading, mathematics, and science results"
[68] This sampling process should be equal to the NAEP in consistency with "relevant, widely accepted professional assessment standards." (*NAEP Authorization Act, SEC 303 (b) (2) (A))*

The sampling process for the NAEP is structured to produce data that is representative of national and statewide performance, so for any given data set the process begins by defining the population (i.e. the US national subset, versus Texas state sample, versus California state sample). According to the NCES, "probability samples[69] of schools and students are selected to represent the diverse student population in the United States. A national sample will have sufficient schools and students to yield data for public schools, each of the four NAEP regions of the country, as well as sex, race, degree of urbanization of school location, parent education, and participation in the National School Lunch Program (NSLP)." Stratification between schools, with random samples drawn from strata based on type, location, and size of schools, allows for additional differentiation in the process.[70] In 2002, the NAEP sampling process further reduced the "burden of testing" by combining the sample drawn for state and national representation: state samples are identified to make state estimates, and from these samples of all 50 states, a sub-sample of students is drawn for the national subset.[71]

The randomness of the selection process and subsequent participation of selected students ensures that "the selection of each student is unrelated to any feature of property of that student," which in turn enables the NAEP's representativeness on national and state scales (Kolstad). This equal representation enables inferences about student achievement across the nation and within individual states and regions, with further demarcation among demographics (in age, gender, race, socioeconomic status, etcetera). EdSource, a non-profit source on issues in education,

[69] "A sample in which every element of the population (the entire collection of American students in public or private schools at grades 4, 8, or 12 [or in the case of the long-term trend assessments, at ages 9, 13, and 17 years]) has a known, nonzero probability of being selected." National Center for Education Statistics (via online webpage)

[70] Sampling weights are employed to prevent occurrence of bias that may be introduced by variations in size for each stratum.

[71] *How the Samples of Schools and Students Are Selected for the Main Assessments (State and National),* National Center for Education Statistics (via online webpage)

reports that the data produced by the sampling process "allow[s] complete coverage of the subject being assessed" and "produces accurate estimates of student achievement," while minimizing time administering the exam, costs associated with assessment, and students participating in assessment procedures.[72] For example, NCES reports that out of 4.1 million fourth–grade students and 3.8 million eighth-grade students, NAEP only tests 180,000 in each grade level; and out of 2.6 million twelfth–grade students, NAEP only tests 16,000 (Kolstad). Yet, the sampling procedure utilized in the NAEP has not limited its utility in research or policy spheres: "The information serves many purposes for a broad constellation of audiences, including researchers, policymakers, the press, and the public. These audiences, both the more technical users and the lay public, look to NAEP to support, refute, or inform their ideas about the academic accomplishments of students in the United States " (*NAEP Reporting Practices,* 87).

Similarly, student performance data produced by a statewide system of stratified random sampling can reliably and validly inform State accountability systems. Scaling the populations sampled from nation-level to state-level, and state-level to district-level, while bolstering assessment participation regulations and establishing meta-accountability for the assessment itself,[73] enables implementation of State PR 1 to replace current assessment methods for statewide accountability. Federal PR 1, which recommended lowering stakes upheld by ESSA regulations, facilitates implementation of State PR 1, by eliminating potentially preventative federal regulations, such as the requirement for 95% (minimum) student participation in state standardized assessments. By producing valid and reliable data that is representative across the state and differentiable across districts, schools, and demographics, State PR 1's recommendation

[72] EdSource.org, "Frequently Asked Questions About NAEP Sampling: Developed in support of reporting 2005 NAEP reading, mathematics, and science results"

[73] At the national level, NAEP assessment and sampling procedures are critically assessed by the NAEP Validity Studies Panel (NVSP), formed by the American Institutes for Research under contract with NCES

to utilize stratified random sampling maintains a standardized system for development of informative data at state level (EPP 3). By minimizing the number of students tested annually and ensuring students are randomly selected for assessment (and thus vary from year to year), State PR 1 lowers stakes tied to classroom performance on standardized assessment (EPP 4). By presenting a method for retaining statewide standardized assessments without the harmful iatrogenic effects, State PR 1 enables an additional indicator in a multiple indicator accountability system (EPP 5).

iii. State PR 2: Incorporate Demographics in Need-Based Formula for School Comparison

Figure 11: State Policy Recommendation 2 (State PR 2)

Policy Level	Recommendation	EPP's Addressed	Means for Satisfying EPP
State	Need-based evaluation of schools (consideration of demographics) for funding purposes	EPP 1	Ensures that need is accounted for in determining systems that allocate resources and funding, reducing the risk of cyclical school failure

Previous chapters discussed the negative unintentional consequences of merit-based, rather than need-based, systems of education. EPP 1 in chapter two presents evidence justifying the principle that basing systems of accountability on merit, rather than need, has unintentional detrimental consequences.[74] Federal PR 1 in chapter four examines trends in student performance data that demonstrate the achievement gap. In particular, these trends indicate that

[74] See chapter II, section ii: "Discussion of Education Policy Principles"

education policies prioritizing need and equality (under ESEA as it was originally passed) narrowed the achievement gap, but when policies shifted priority to merit (under "excellence" reforms, such as NCLB) that trend reversed.[75] These trends indicate the potential for need-based reforms to enable closure of the achievement gap. Thus, State PR 2 prescribes re-prioritizing need in policies at the state level. Specifically, this section proposes to do this by incorporating demographics – both income and race – into the formula for determining distribution of aid funding to schools.

As previously discussed, Robenstine argues that merit is not a viable basis for systems of education without first considering issues of equity, asserting that, "the justification of meritocracy as a dominant, controlling assumption is valid *if and only if* equal educational opportunity exists first" (Robenstine, 177). The persistence of the achievement gap (and other inequities), then, necessarily shifts the focus for current reforms back toward equity issues, and equity demands observance of need. Methods used by states both to identify schools receiving aid funding and to determine proportional allocations of that aid to qualifying schools should thus recognize concentration of high-need demographics in the formula for these determinations. By allotting weight relative to the density of low-income and minority students, this formula will re-prioritize need and justify the consideration of merit. An example for funding allocations for four schools – two high need, low performing schools; two low need, mid-to-high performing schools – is presented on the following page in Figure 12. The need-based formula used for this example prioritizes need rather than merit, allocating 70% of the total funds on a need-basis, and allocating 30% of the total funds on merit-basis. Figure 13 displays the distribution process under this formula for one school in particular, School A. *[Refer to Figures 12 and 13 on the following page]*

[75] See chapter IV, section ii: "Policy Recommendation 1: Reauthorize ESEA with Lower Stakes"

Figure 12: Funding Allocations in a Hypothetical Need-Based Formula for Aid

The following table exhibits funding allocations under a need-based system of aid for four schools: Schools A, B, C and D. Schools A and B are Title I, high need, low performing schools; Schools C and D are low-need, mid to high performing schools. While merit is considered, need is weighted significantly heavier. In this example, need is given 70% weight, and merit is given 30% weight. The Need Score (N) earned by a school should depend on factors such as concentration of low-income and minority students, existing resources, and total number of students; in this example, a school may earn up to 10 points for N. The Merit Score (M) earned by a school should depend on factors such as graduation rates, teacher retention and satisfaction, and student performance on statewide assessments or performance-based assessments.

	School A*	School B*	School C	School D	Total
Need Score (N)	8	10	1	3	22
Percentile (N÷22)	36.4%	45.5%	4.5%	13.6%	100%
Need-Based Award	**$25.45**	**$31.82**	**$3.18**	**$9.55**	**$70.00**
Merit Score (M)	3	1	8	6	18
Percentile (M÷18)	16.7%	5.6%	44.4%	33.3%	100%
Merit-Based Award	**$5.00**	**$1.67**	**$13.33**	**$10.00**	**$30.00**
Total Funding Awarded	**$30.45**	**$33.49**	**$16.51**	**$19.55**	**$100.00**

* Denotes Title I School

Weight distribution: 70% Need; 30% Merit

T = Total funding available to district*
F = Total funding awarded to school*
A^N = Need-Based Award*
A^M = Merit-Based Award*
S^N = Sum of N for all schools
S^M = Sum of M for all schools

N = Need Score
(Scored from 1-10, based on existing resources and number of low-income and minority students)
M = Merit Score
(Scored from 1-10, based on student performance on statewide or performance-based assessments)

**(In thousands of dollars)*

Funding Allocated for School A
(High need, low performing)

T = $100

Need: 70% = $70

N = 8

$\frac{N}{S^N} = \frac{8}{22} = 36.4\%$

36.4% of $70

A^N = $25.45

Merit: 30% = $30

M = 3

$\frac{N}{S^M} = \frac{3}{18} = 16.7\%$

16.7% of $30

A^M = $5.00

$F = A^N + A^M$ = $25.45 + $5.00 = **$25.45**

Figure 13: Formulaic Procedure to Determine School A Funding Allocation

The table at the left demonstrates the process used to determine total funding awarded to schools (F) under a need-based system of aid. The table determines the funds allocated to School A, a high need, low performing school, under the need-based formula used in this example. Because heavy weight (70%) is allocated to need, School A is awarded the second largest allocation of funding, and displays the second highest need. Figure 8, and the key below Figure 8, may be used to reference the letters and numbers used in this formula.

In terms of distribution of aid funding, discussions on the issue of merit versus need often occur at the post-secondary level of education. Most of these arguments naturally center on the expansion of post-secondary attendance opportunities for low income students; this is not applicable at the primary and secondary levels in public education, because opportunity for attendance is generally not limited by income level.[76] However, the efficiency argument in particular is equally applicable in public K-12 education. Because prioritizing need streamlines funds directly to only the schools in which resources are most urgently needed (Hoenack in Doyle, 399), "[t]his efficiency means that [need-based] programs are more affordable for the state as a whole than other, more broad-based subsidies" (Hearn and Longanecker in Doyle, 399). Similarly, by employing the proposed need-based formula, such as that used in Figures 12 and 13, State Education Agencies (SEAs) would more efficiently allocate aid to district schools by concentrating funds within schools with highest need.

The objective of a merit-based system, to incentivize and subsequently stimulate achievement, is not in itself ill intended. The issue, as Robenstine has contended, is that these objectives are unattainable in a system that maintains unequal access to educational opportunities – an issue still very much present in the American education system (Orfield and Lee, 5). Instead, merit thus serves to bolster these inequities by best serving those students who have the most access to opportunities in education, and most harming those with the least access. Given that the students with the least opportunities tend to belong disproportionately to low-income and minority demographics, merit-based systems reinforce, rather than interrupt, the marginalization of these demographics (Robenstine, 117). The flaws of such a system are

[76] This statement holds in the general sense. Arguments could be made, however, that attendance is limited by income level, when considerations are made regarding (a) student transportation and (b) student employment in the workforce to earn supplemental household income.

similarly evident at the school level: those without the resources to provide opportunities to students cannot generally, for this very reason, outperform competitors in a meritocracy. The meritocracy, at that point, excludes certain schools from the 'race.' It is clear, then, that 'achievement' reforms do not generate equity. Thus, equity, rather than merit, must inform the process of allocating funding and resources that facilitate opportunities for students in schools. This proposal to incorporate and heavily weigh need in the determination of funding allocations addresses the issues of merit that were discussed initially in EPP 1, progressing toward a more equitable system of education.

The following chapter introduces and defends a recommendation for education policy at the local level. This recommendation focuses on shifting assessments in from measuring rote skills to developing higher-order learning and twenty first century skills. Lower stakes at the federal level, in addition to restructured accountability and a paradigm shift at the state level, facilitate local use of a performance-based system of assessment.

VII. Recommendation for Policy at the Local Level

As with the policy recommendations at the state level, recommendations for policy at the local level presented in this chapter do not center on a specific local region; rather, this recommendation is proposed as a policy "ideal," for implementation within state and federal jurisdictions that provide the according flexibility for this local recommendation. Thus, this PR serves as a general model for one reform at the local level: the previous recommendations at the federal level aim to lower the stakes tied to accountability, the recommendations at the state level shift paradigms for accountability, and this recommendation at the local level proposes an alternative form of assessment for use within a new statewide accountability system. The Local PR recommends using a system of performance-based assessments to measure student performance. This PR addresses EPP 2 by enabling and promoting development of twenty first century skills and EPP 5 by acting as an additional indicator in an otherwise single indicator system of performance measurement.

i. The Local Role in Education Policy

The local role in education policy is primarily concentrated in the responsibilities belonging to Local Education Agencies (LEAs). LEAs generally function at the city or district level, and, more so than any other level of government, they directly impact schools, teachers, and students. As defined by the US Department of Education, LEAs serve the political subdivisions of states (such as districts), as "a public board of education or other public authority legally constituted within a State for either administrative control or direction of, or to perform a

service function for, public elementary schools or secondary schools."[77] While LEAs act as the narrowest structural governing institution in education apart from schools themselves, they hold considerable political power in their authority to implement both federal laws, such as ESSA and IDEA regulations, and state laws, such as accountability assessments. Further, LEAs generally shape the specific practices that adhere to those broader policies. Because of this significant authority belonging to LEAs, the US Department of Education calls the local level, "the heart of the U.S. education system" for primary and secondary education, enlisting their specific power to "operate schools, implement and enforce state laws and policies, develop and implement their own educational policies, hire and supervise professional teaching staffs, and raise money to pay for schools (usually through property taxes plus special bond issues)."[78]

The governing body of school districts takes the form of school boards, comprised of elected citizens who have broad powers of oversight. While these powers are significant, they are primarily rooted in *oversight* of state and federal regulations. LEAs act in their role to identify and group schools in terms of performance. State programs collect this grouping information and related data, typically to inform reports such as State Performance Plan (SPP) or the Annual Performance Report (APR). Perhaps more so, the data for performance groupings inform the organization of local initiatives. Through this capacity, local-level programs may emerge: In New York City, for example, the mayor has a significant role in local policy, creating specific programs under which local schools are chosen to be grouped. Mayor Bill de Blasio's largest action in this role has been in creating the Renewal School Program, a program that identifies low-performing schools (currently 86 in NYC), and equips those schools with resources and

[77] 20 USCS § 7801(26)(A)
[78] "Organization of U.S. Education: The Local Role," *International Affairs Office, U.S. Department of Education,* Feb. 2008.

support, with the intention of stimulating short-term improvement turnaround.[79] The program has received mixed criticism, however. Kate Taylor, in an article for the New York Times, attributes this to the high-stakes nature of the program, and the ensuing "pressure" to perform on standardized assessments: "no school wants to be on the chopping block."[80] Thus, local-level policies are not immune to policies that raise stakes for standardized assessments, even if those assessments and the primary level of accountability for performance on those assessments rests at the state level. The scale of authority resting at the local level is reflected in its sourcing of funds to schools. According to data reported by NCES Common Core of Data (CCD), Local sources of funding accounted for 46% of total funding in 2012-2013, the largest portion among state and federal sources of funding and "the highest percentage in the past 10 years."[81] This portion is near equal to that accounted for by state funding (45%) in the same year, and significantly higher than federal funding (13%).

The cornerstone of the LEA role, however, is its proximity to schools. This, in addition to the local authority to allocate funds and oversee federal and state regulations, furnishes the local level with a significant role in influencing the practices within schools and classrooms. Consequently, this role can manifest in decidedly impactful ways. For example, the Renewal School program in New York City is a partnership between NYC Department of Education and the local school leaders to turn around the performance of 86 struggling schools in New York. The program demonstrates the local authority to directly impact school practices. Thus, this

[79] NYC Department of Education Online Web Page, *Renewal Schools* (http://schools.nyc.gov/AboutUs/schools/RenewalSchools/default)
[80] Taylor, Kate. "Pressure Builds for City's Renewal Schools as State Tests Begin." *New York Times.* March 28, 2017. (https://www.nytimes.com/2017/03/28/nyregion/new-york-city-renewal-schools-test-prep.html?_r=0)
[81] Musu-Gillette, Lauren and Stephen Cornman, "Financing education: National, state, and local funding and spending for public schools in 2013." *NCES Blog, National Center for Education Statistics.* Jan. 25, 2016. (https://nces.ed.gov/blogs/nces/post/financing-education-national-state-and-local-funding-and-spending-for-public-schools-in-2013)

Local PR recommends utilizing that authority to alter the methods of assessment used to determine student performance. Federal PR 1's recommendation to lower stakes and ease guidelines associated with ESSA will facilitate state-level mandates for a new, lower stakes system of accountability. This, in turn, provides the necessary flexibility for the alternative system of assessment proposed in this Local PR, and – perhaps more importantly – ensures that its measurements of student performance can factor into statewide accountability systems. Employing a new system of assessment at the local level will thus simultaneously rely on and inform a *new* accountability at the state level.

ii. Local PR: Introduce Performance-Based Assessments as Performance Indicator

Figure 14: Local Policy Recommendation (Local PR)

Policy Level	Recommendation	EPP's Addressed	Means for satisfying EPP
Local	Performance Based Assessments	EPP 2	Enables evaluation of skills on an individual level, promotes student and teacher autonomy, and develops twenty first century learning skills
		EPP 5	Constitutes an additional indicator of student performance in a multiple indicator accountability system

Chapter I explores the charge, repeated by education researchers and policy analysts, that high-stakes statewide standardized assessments have unintentional consequences for students, and ultimately concludes that risks associated with these consequences currently outweigh the

benefits of continuing such a system. This Local PR advocates for district-wide utilization of stratified random sampling techniques, in order to maintain the reliable student performance data while minimizing the number of students taking statewide-standardized assessments. Though institutions of education must be held accountable for the performance of *every* student, it is not necessary that this performance be on statewide standardized assessments. While this recommendation adheres to several of the Education Policy Principles discussed in chapters II and III to mitigate consequences of an accountability system based on performance on statewide standardized assessments, it does not address charges that standardized assessments do not accurately measure or promote better instruction or learning. Research consistently finds that the testing mechanisms utilized by standardized assessments can breed environments of rote learning in classrooms:

> However, the main purpose of standardized testing is to sort large numbers of students in as efficient a manner as possible. This limited goal, quite naturally, gives rise to short-answer, multiple-choice questions. When tests are constricted in this manner, active skills, such as writing, speaking, acting, drawing, constructing, repairing, or any of a number of other skills that can and should be taught in schools are automatically relegated to second class status. (Bowers; as cited in Sacks, 28).

So, while State PR 1 addresses data measurement inherent to *systems of accountability*, the remaining issue to address becomes the mechanism for *accountability itself*. Accordingly, this Local PR prescribes an additional indicator of student performance: a non-standardized[82] performance-based system of alternative assessment.

Alternative "performance-based assessment," for the purposes of this policy recommendation, refers to a method of assessment in which students perform a task in order to

[82] "Non-standardized" does not assume inherent differences in method of assessment. While the performance basis of the assessment means that students may produce different work, for utilization in systems of accountability, that student work must be assessed in a standardized manner across the state. A practical means to achieve this is by statewide-standardized rubrics.

demonstrate both content knowledge and satisfactory skills in the subject or unit tested. The tasks in which student performance is assessed can range from analytical essays reviewing a significant literary work, to a student-led science experiment, to mathematical analysis and application toward solving "real-world" problems. Researchers and educators alike find performance-based assessments effective in facilitating student learning and achievement, and appear to prefer them to high-stakes standardized assessments (SRN, 2; Flynn, 33; Parker and Gerber, 66). In application of performance-based assessments during a five-week summer enrichment program, in which preservice teachers assessed the performance of nine low-income, below-average performing African American students in a science exhibit, evaluators found that, "Performance-based assessment appears to be a viable approach for measuring students' knowledge and skills . . . [and] data from individual evaluations and group-negotiated evaluations of performance-based assessment were found to be effective in measuring students' knowledge and skills" (Parker and Gerber, 66). Stanford Redesign Network (SRN) in the Stanford University School of Education reports that performance-based assessments are increasingly necessitated by the evolving demands of the US workforce, which have shifted from "fact-oriented curriculum to one that emphasizes problem solving and innovation" (Herman 1992; as cited in SRN, 2). Further, SRN reports that performance-based assessments correspond to manners of learning better than their counterpart standardized assessments:

> "Performance-based assessment requires students to use high-level thinking to perform, create,or produce something with transferable real-world application. Research has shown that such assessment provides useful information about student performance to students, parents, teachers, principals, and policymakers. Research on thinking and learning processes also shows that performance- based assessment propels the education system in a direction that corresponds with how individuals actually learn." (SRN, 2)

Educators, too, appear to prefer the method. In a 2008 write-up of her outlook on and experience with evaluating students using performance-based assessments, a 20-year veteran teacher (self described as "jaded" by the high-stakes testing movement) wrote that she was "pleasantly surprised" both with her students exceedingly high scores on the assessments, and "with [her] own rejuvenated interest in testing, as well" (Flynn, 35). The benefits for students in, and the utility of, a performance-based system of assessment are further evidenced in the continued success of the New York Performance Standards Consortium (NYPSC).

NYPSC, originally formed in 1997 under a waiver of exemption from the New York Regents exam, is a consortium of (currently) 38 public non-charter high schools in New York, 36 of which are in New York City, that assess students using an alternative to high-stakes standardized assessments in the form of four required Performance-Based Assessment Tasks (PBATs): "an analytic literature essay, a social studies research paper, a student-designed science experiment, and higher-level mathematics problems that have real-world applications. They include both written and oral components" (FairTest, 2014). Educators look to the Consortium's success in using PBATs as a hopeful indicator for future accountability systems: Michelle Fine, Psychology and Urban Education professor at the City University of New York Graduate Center, wrote that in response to NYPSC data, "[she] thought so *this* is what accountability should look like: a model of *complex* accountability." Consortium schools serve a diverse range of students and demographics that align proportionally to those within the NY public school system, and data reflects the success of NYPSC compared to New York public schools.

Consortium schools outperform NY public schools, in addition to averages for schools across the nation, on nearly every front. Consortium schools had equal or slightly greater numbers of black and Hispanic, English language learning, special needs, and low income

students than NYC high schools, yet Consortium schools outperformed NYC high schools 68.6% to 59% in 4-year graduation rates[83] for the 2008-2009 year, and in persistence in college, with students graduating from Consortium schools continuing for a second year in both 4-year and 2-year universities at significantly higher rates than those for national and NYC high school averages in 2008.[84] Consortium schools also exceedingly outperformed national averages for minority male acceptance rates, with 86% of African-American and 90% of Latino males accepted to college (compared to 37% African-American and 42% Latino males, nationally). The success extends beyond student academics: Consortium schools had drastically lower teacher turnover rates than NYC high schools (15% compared to 58% turnover in NYC schools) and suspension rates (5% compared to 12% suspension in NYC schools). FairTest reports in the NYPSC fact sheet that, "Performance-based assessment works well for all students, but [NYPSC's] success with the most vulnerable students is what makes [the outcomes] impressive" (FairTest, 2016).

The noteworthy success of the NYPSC, in both citywide average and nationwide average comparisons, demonstrates the potential for utilization of performance-based assessments for complex statewide accountability systems. In particular, these numbers suggest that considerable benefits are produced by performance-based assessments, in terms of graduation, college acceptance, and college persistence rates, for minority and low-income students. By

[83] Outperformance in graduation rates held for 5-year graduation rates (76% in Consortium schools to 66.1% in NYC high schools), and across differentiated demographics: 60.8% (Consortium) to 53.9% (NYC) among black students, 64.9% to 51.8% (Hispanic), 87.6% to 76.8% (Asian), 77.9% to 73.9% (white), 69.5% to 39.7% (ELL), 50% to 24.7% (special needs). These numbers indicate that Consortium focus on PBAT's was significantly beneficial to *all subsets* of students.

[84] 4-year college 2nd year persistence: 93.3% (Consortium) versus 74.7% (National) and 80.8% (NY schools); 2-year college 2nd year persistence: 83.9% (Consortium) versus 53.5% (National) and 59.1% (NY schools). (Via "Chart 2: Persistence in College: 2nd Year Comparison Between Consortium, National and NYS Rates, Class of 2008" in *Educating for the 21st Century*)

simultaneously assessing content mastery and performance skills, performance-based assessments in systems of accountability will promote, rather than hinder, higher-order learning that fosters 21st century skills (EPP 2).

VIII. Conclusion

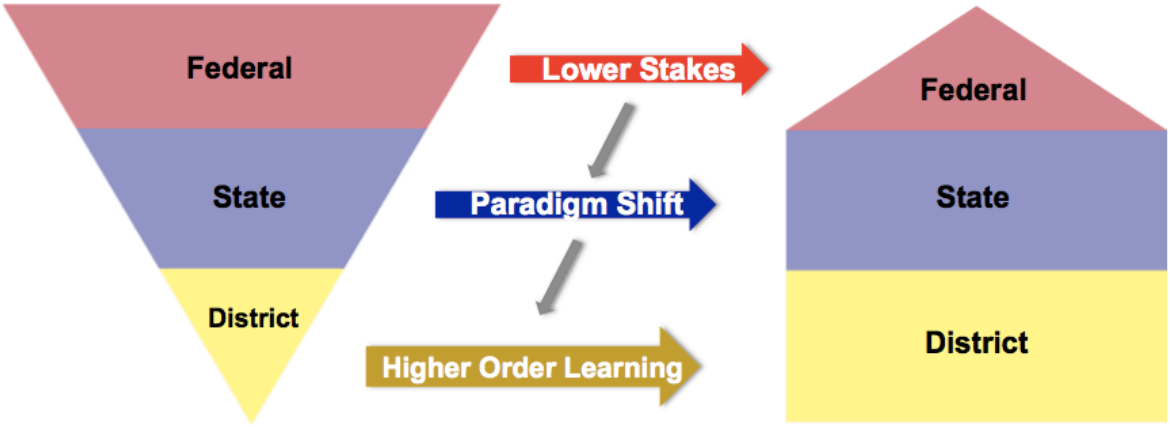
If the heart of education is the student experience in the classroom, then teachers and schools keep the heart beating. Policy at the federal, state and local levels must provide the structure to ensure that the process continues – equitably and satisfactorily. Unfortunately, the United States has a long history of inequity. Nearly sixty years ago, Brown v. Board of Education (1954) ruled against legal segregation in schools, beginning the extensive road toward toward equity in the U.S. public school system. The persistence of a wide achievement gap (Guisbond, et. al, 2013) is evidence of the remaining issues policy at all levels of government must rise to tackle. Failures in progress toward equal opportunity, though, are not for lack of trying, particularly at the federal level. This paper discusses many of these policy measures in-depth: from the Elementary and Secondary Education Act (ESEA) of 1965, to No Child Left Behind (NCLB) in 2001, to the Every Student Succeeds Act (ESSA) in 2015. While ESEA's creation of Title I aid funding for low-income schools was a powerful step toward equity, later attempts to stimulate achievement in NCLB entrenched the accountability system with stakes at the federal level. Analysis of ESSA – though it returns the standards of accountability to state authority – indicates that its retention of stringent guidelines for accountability may lead to similar, unintended consequences that appeared as a result of NCLB: a proliferation of high stakes assessments that measure rote skills rather than higher order learning. This thesis aimed to redirect policy focus away from the stakes and lower-order testing of these past reforms. An investigation of the effects of education policies pointed to five principles for future reform. Research to determine the means for best implementing these reforms in policy, however, pointed to differing levels of authority in a decentralized government. This revealed the necessity to rethink reform and accountability at the *intersection* of federal, state and local policy.

This discovery may potentially shed light on the consequences of past reforms. Broad, sweeping measures at the top – such as federally mandated accountability under NCLB – do not provide the flexibility at state and local levels for the needs of every school, classroom, and student to be met. This is evident in states' eventual reliance on waivers under the Obama administration to gain exemption from certain regulations. Similar concerns may be raised regarding discussions of future reforms centered on federally mandated school choice. Critics argue that choice does not work in every district, in every state (Carr 2017, Prothero 2017).

This thesis consequently culminated in a broad evaluation of policy measures, which were subsequently coordinated across the varying levels of government to indicate the five Policy Recommendations (PRs) outlined in the body of the thesis.[85] The synthesis of these recommendations revealed broader themes that must be addressed at the various levels of government, simultaneously and cohesively. First, individual policies were identified through varying methods of research to be imperative steps toward achieving equity and quality in education. These policies (the PRs), when consolidated and scaled to broader themes (see Figure 15, below) indicated the necessary actions that must be taken at each level. Federal PRs 1 and 2 indicated the need to lower stakes at the federal level, State PRs 1 and 2 indicated the need to shift paradigms from high stakes and merit to equity and need at the state level, and Local PR 1 indicated the need to facilitate higher-order learning in accountability assessments at the local level. Simultaneously, this action at the federal level facilitates the necessary actions at the state and local levels. Thus, an intersectional approach to reform is not only beneficial, but necessary.

[85] The five PRs were: Reauthorize ESEA with lower stakes (Federal PR 1), eliminate and replace RTTT with a grant that incentivizes intelligent accountability (Federal PR 2), utilize stratified random sampling to measure student performance on statewide standardized assessments (State PR 1), incorporate need-based demographics in the formula for determining allocation of school funding (State PR 2), and implement a system of performance-based assessments to use for accountability purposes (Local PR).

Figure 15: Concentration of Power and Necessary Shifts
The figure below depicts the course of action at each level of government, based on broader themes indicated by the individual PRs. The structure of the shapes reflects the concentration of stakes set at each level tied to the past, current, and (on the right) new systems of accountability. Note that the broad thematic actions at each level of policy are connected by a gray arrow, which indicates the intersectional reliance of the policies in order to achieve successful implementation.



The significance of the broader implications of the Education Policy Principles (EPPs) and the PRs is that the policies recommended in this thesis are not the golden ticket to a fully equitable and quality education system. *Many* other policies must accompany the ones recommended in the previous other chapters. Due to limitations of scope, resources and time – this thesis cannot and does not claim to "fix" the United States education system. As Figure 15 demonstrates, however, there are *themes for action* that should guide future reforms. The federal, state and local Policy Recommendations provide an introductory path toward a new accountability system – accountability with the flexibility and coordination to accommodate the growth of *all students.*

Bibliography

Advancement Project, Education Law Center - PA, FairTest, The Forum for Education and Democracy, Juvenile Law Center, and NAACP Legal Defense and Educational Fund, Inc., comps. *Federal Policy, ESEA Reauthorization, and the School-to-Prison Pipeline*. N.p., 2011.

Anrig, Gregory. "Educational Standards, Testing, and Equity." *The Phi Delta Kappan* 66, no. 9 (May 1985): 623-25.

Anson, Chris M. "Closed Systems and Standardized Writing Tests." *College Composition and Communication* 60, no. 1 (September 2008): 113-28.

Armendariz Maxwell, Cynthia. Changing Student and Teacher Demographics within the Current High-stakes, Standards-based, Assessment-driven Atmosphere of American Public Education: Mentoring Educators of English Language Learners. Diss., 2009. Ann Arbor: ProQuest, 2010. Print.

Baek, Sun-Geun. "Implications of Cognitive Psychology for Educational Testing." *Educational Psychology Review* 6, no. 4 (December 1994): 373-89. http://www.jstor.org/stable/23359295.

Baines, Lawrence, and Gregory Kent Stanley. "The Iatrogenic Consequences of Standards-Based Education." *The Clearing House* 79, no. 3 (January/February 2006): 119-23.

Bongiorno, Deborah, ed. *Using Student Achievement Data to Support Instructional Decision Making*. Illustrated by Donna Sicklesmith-Anderson. N.p.: National Association of Elementary School Principals, 2011.

Booher-Jennings, Jennifer. "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal* 42, no. 2 (Summer 2005): 231-68.

Bovaird, James A., Kurt F. Geisinger, and Chad W. Buckendahl. High-stakes Testing in Education: Science and Practice in K-12 Settings. 1st ed. Washington, DC: American Psychological Association, 2011. Print.

Brown, C. P. "Children of Reform: The Impact of High-Stakes Education Reform on Preservice Teachers." Journal of Teacher Education 61.5 (2010): 477-91. Web.

Brown, Stephanie N. "A New Era of Educational Assessment: The Use of Stratified Random Sampling in High Stakes Testing." PhD diss., University of North Texas, 2013.

Burrus, Jeremy, Teresa Jackson, Nuo Xi, and Jonathan Steinberg. *Identifying the Most Important 21st Century Workforce Competencies: An Analysis of the Occupational Information Network (O*NET)*. N.p.: Educational Testing Service, 2013.

Bussert-Webb, Kathy. "To Test or to Teach: Reflections from a Holistic Teacher-Researcher in South Texas." *Journal of Adolescent & Adult Literacy* 42, no. 7 (April 1999): 582-85. http://www.jstor.org/stable/40015640.

Camera, Lauren. Lauren Camera to U.S. News and World Report newsgroup, "Achievement Gap Between White and Black Students Still Gaping," January 13, 2016. https://www.usnews.com/news/blogs/data-mine/2016/01/13/achievement-gap-between-white-and-black-students-still-gaping.

Carr, Sarah. "Betsy DeVos' Big Education Idea Doesn't Work." *Slate*. http://www.slate.com/articles/news_and_politics/schooled/2017/01/betsy_devos_big_education_idea_doesn_t_work.html.

Center for Research on Evaluation, Standards and Student Testing University of Colorado at Boulder. *Complex Performance Based Assessment: Expectations and Validation Criteria*. By Robert Linn, Eva Baker, and Stephen Dunbar. N.p., 1990.

Coleman, Arthur L. "Excellence and Equity in Education: High Standards for High Stakes Tests." *Virginia Journal of Social Policy & the Law*, 1999, 81-113.

Coleman, Arthur L. "Excellence and Equity in Education: High Standards for High Stakes Tests." Virginia Journal of Social Policy & the Law, 1999, 81-113.

Coleman, Arthur L. "Excellence and Equity in Education: High Standards for High-Stakes Tests." Virginia Journal of Social Policy & the Law 6.1 (1998): 81-113. University of Virginia School of Law. Web. 05 Nov. 2016.

Collier, Kiah. "Lawmakers Look at Tying School Funding to Performance." *The Texas Tribune*. Accessed August 3, 2016. https://www.texastribune.org/2016/08/03/senators-examining-performance-based-funding-schoo/.

Corporate Profit at Equity's Expense: Codified Standards and High-Stakes Assessment in Music Teacher Preparation (Julia Eklund Koza)

Dalton, Ben, Elizabeth Glennie, Steven J. Ingels, and John Wirt. *Late High School Dropouts: Characteristics, Experiences, and Changes Across Cohorts*. N.p.: National Center for Education Statistics, 2009.

Darling-Hammond, Linda. "Transforming Urban Public Schools: The Role of Standards and Accountability." *ERIC*, 2000.

Davey, Lynn. "The Case for a National Testing System. ERIC Digest." *ERIC*. Last modified 1992. http://ericdigests.org.

Doyle, W.R. "Does Merit-Based Aid 'Crowd Out' Need-Based Aid?" *Research in Higher Education* 51, no. 5 (August 2010): 397-415.

Ellison, Scott. "Intelligent Accountability: Re-Thinking the Concept of "Accountability" in the Popular Discourse of Education Policy." *Journal of Thought* 47, no. 2 (Summer 2012): 19-41. http://www.jstor.org/stable/jthought.47.2.19.

———. "Equality and Education: Fifteen Years On." *Oxford Review of Education* 17, no. 2 (1991): 161-67.

Eyesenck, H.J. "Equality and Education: Fact and Fiction." *Oxford Review of Education* 1, no. 1 (1975): 51-58.

Fahy, Colleen. "Education Funding in Massachusetts: The Effects of Aid Modifications on Vertical and Horizontal Equity." *Journal of Education Finance* 36, no. 3 (Winter 2011): 217-43.

FairTest, comp. *Racial Justice and Standardized Educational Testing*.

Faria, Ann-Marie, Jessica Heppen, Suzanne Stachel, Wehmah Jones, Katherine Sawyer, Kerri Thomsen, Melissa Kutner, David Miser, Sharon Lewis, Michael Casserly, Candace Simon, Renata Uzzell, Amanda Corcoran, and Moses Palacios. "Charting Success: Data Use and Student Achievement in Urban Schools." *American Institutes for Research*, Summer 2012.

Farrell, Patricia, and Gregory Kienzl. "Are State Non-Need, Merit-Based Scholarship Programs Impacting College Enrollment?" *Education Finance and Policy* 4, no. 2 (Spring 2009): 150-74.

———. "The Federal Role in Education." Last modified July 21, 2016. https://www2.ed.gov/about/overview/fed/role.html.

Flynn, Lee-Ann. "In Praise of Performance-Based Assessments." *Science and Children* 45, no. 8 (April/May 2008): 32-35.

Foorman, Barbara, Sharon Kalinowski, and Wayne Sexton. Standards-Based Educational Reform Is One Important Step Toward Reducing the Achievement Gap to *Standards-Based Reform and the Poverty Gap*, by Adam Gamoran. N.p.: Brookings Institution Press, 2007.

Foote, Martha. *New York Performance Standards Consortium: ??College Performance Study*. N.p., 2005.

Gagnon, Paul. "The Case for Standards: Equity and Competence." *The Journal of Education* 176, no. 3 (1994): 1-16.

Garrison, Mark J. A Measure of Failure: The Political Origins of Standardized Testing. Albany: State U of New York, 2009. Print.

Goldberg, Milton, Susan Traiman, Alex Molnar, and John Stevens. "Why Business Backs Education Standards." *Brookings Papers on Education Policy*, no. 4 (2001): 75-129.

Gomez-Velez, Natalie. "Urban Public Education Reform: Governance, Accountability, Outsourcing." *The Urban Lawyer* 45, no. 1 (Winter 2013): 51-104.

González-Lloret, Marta. *A Practical Guide to Integrating Technology into Task-Based Language Teaching*. N.p.: Georgetown University Press, 2015.

Gorlewski, Julie, Brad Porfilio, and David Gorlewski. Using Standards and High-Stakes Testing for Students: Exploiting Power with Critical Pedagogy. Counterpoints: Studies in the Postmodern Theory of Education. Vol. 425. New York: Peter Lang, 2012. Print.

Griffin, Des. Education Reform: The Unwinding of Intelligence and Creativity. Cham: Springer, 2014. Print.

Guisbond, Lisa, Monty Neill, and Bob Schaeffer. "NCLB's Lost Decade for Educational Progress: What Can We Learn from This Policy Failure?" *Counterpoints* 451 (2013): 7-26.

Hargraves, Andy, Lorna Earl, and Michele Schmidt. "Perspectives on Alternative Assessment Reform." *American Educational Research Journal* 39, no. 1 (Spring 2002): 69-95. http://www.jstor.org/stable/3202471.

Harris, Elizabeth. "20% of New York State Students Opted Out of Standardized Tests This Year." *The New York Times*. Accessed August 12, 2015. https://www.nytimes.com/2015/08/13/nyregion/new-york-state-students-standardized-tests.html.

Hout, Michael, and Stuart Elliott, eds. *Incentives and Test-Based Accountability in Education*. Washington, D.C.: The National Academies Press, 2011.

Howe, Kenneth. "Standards, Assessment, and Equality of Educational Opportunity." *Educational Researcher* 23, no. 8 (November 1994): 27-33.

Hursh, David W. High-stakes Testing and the Decline of Teaching and Learning: The Real Crisis in Education. Lanham: Rowman & Littlefield, 2008. Print.

International Affairs Office, and U.S. Department of Education. *Organization of U.S. Education: State Role I - Primary and Secondary Education*. http://www.ed.gov/international/usnei/edlite-index.html.

Jennings, Jack. *Presidents, Congress and the Public Schools: The Politics of Education Reform*. Cambridge, MA: Harvard Education Press, 2015.

Jochim, Ashley, and Patrick Murphy. *The Capacity Challenge: What it Takes for State Education Agencies to Support School Improvement*. N.p.: Center on Reinventing Public Education, 2013.

Jones, Gary. "The Debate over: No-Need Scholarships: Merit Aid Is an Investment for American Leadership." *Change* 16, no. 6 (September 1984): 24-28.

Jones, Ken, and Betty Lou Whitford. "Let Them Eat Tests: High-Stakes Testing and Educational Equity." *Journal of Thought* 37, no. 4 (Winter 2002): 35-49.

Kai Wah Chu, Samuel, Rebecca Reynolds, Nicole Tavares, Michele Notari, and Celina Wing Yi Lee. Twenty-First Century Skills and Global Education Roadmaps to *21st Century Skills Development Through Inquiry-Based Learning*, 17-32.

Kearns, Laura-Lee. "High-stakes Standardized Testing and Marginalized Youth: An Examination of the Impact on Those Who Fail." *Canadian Journal of Education* 34, no. 2 (2011): 112-30.

Kolstad, Andrew. "Basic Sampling Concepts Used in NAEP." In *The Nation's Report Card*. N.p.: National Center for Education Statistics, 2006.

Lee, Jaekyung, and Kenneth Wong. "The Impact of Accountability on Racial and Socioeconomic Equity: Considering Both School Resources and Achievement Outcomes." *American Educational Research Journal* 41, no. 4 (Winter 2004): 797-832.

Lee, Jaekyung. "Is Test-Driven External Accountability Effective? Synthesizing the Evidence from Cross- State Causal-Comparative and Correlational Studies." *Review of Educational Research* 78, no. 3 (September 2008): 608-44.

Levy, Frank, and Richard J. Murnane. "How Computerized Work and Globalization Shape Human Skill Demands." *Industrial Performance Center*, August 2005.

Listen to Us: ?Teacher Views and Voices. N.p.: The Center on Education Policy, 2016.

Loveless, Tom. *The Peculiar Politics of No Child Left Behind*. N.p.: The Brookings Institution, 2006.

Lynch, Matthew. Reimagining Education Reform and Innovation. New York, NY: Peter Lang, 2014. Print.

Madan, Amman. "Sociologising Merit." *Economic and Political Weekly* 42, no. 29 (July 2007): 3044-50.

Madaus, George, and Michael Russell. "Paradoxes of High-Stakes Testing." *The Journal of Education* 190, no. 1/2 (2010/2011): 21-30.

Marker, Perry. "Standards and High Stakes Testing." *Theory & Research in Social Education* 29, no. 2: 357-62. http://dx.doi.org/10.1080/00933104.2001.10505944.

Marker, Perry. Standards and High Stakes Testing: The Dark Side of a Generation of Political, Economic and Social Neglect of Public Education, Theory & Research in Social Education, 2001. 29:2, 357-362, DOI: 10.1080/00933104.2001.10505944

Mathis, William J., and Tina M. Trujillo, eds. Learning from the Federal Market-based Reforms: Lessons for ESSA. N.p.: n.p., n.d. Print.

McDermott, Kathryn A. *Education Governance for the Twenty-First Century: Overcoming the Structural Barriers to School Reform*. Edited by Paul Manna and Patrick McGuinn. N.p.: Brookings Institution Press, n.d.

McDermott, Kathryn. *High Stakes Reform*. N.p.: Georgetown University Press, 2011.

Mintrop, Heinrich, and Gail Sunderman. "Predictable Failure of Federal Sanctions-Driven Accountability for School Improvement And Why We May Retain It Anyway." *Educational Researcher*, 2009.

Moses, Michele S., and Nanna, Michael J., "The Testing Culture and the Persistence of High Stakes Testing Reforms" Education and Culture, Vol. 23, No. 1 (2007), pp. 55-72. Purdue University Press. Print.

———. "Multiple Facets of Inequity in Racial and Ethnic Achievement Gaps." *Peabody Journal of Education* 79, no. 2 (2004): 51-73.

Murphy, Joseph, and Jacob E. Adams. "Reforming America's schools 1980-2000." *Journal of Educational Administration* 36, no. 5: 426-44. http://dx.doi.org/10.1108/09578239810238438.

NAEP Validity Studies: An Agenda for NAEP Validity Research. N.p.: National Center for Education Statistics, 2003.

Nathaniel von der Embse & Ramzi Hasson (2012) Test Anxiety and High-Stakes Test Performance Between School Settings: Implications for Educators, Preventing School Failure: Alternative Education for Children and Youth, 56:3, 180-187, DOI: 10.1080/1045988X.2011.633285

Neill, Monty. "A Child Is Not a Test Score: Assessment as a Civil Rights Issue." *The Next Phase of Election Reform*, Fall 2009, 28-35.

New York Performance Standards Consortium. *Educating for the 21st Century: Data Report on the New York Performance Standards Consortium*.

Nichols, Sharon L., and David C. Berliner. "The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing." *The Great Lakes Center for Education Research & Practice*, March 2005, 1-171.

Orfield, Gary, and Chungmei Lee. "Why Segregation Matters: Poverty and Educational Inequality." *The Civil Rights Project*, January 13, 2005.

———. *Organization of U.S. Education: The Local Role*.
"Frequently Asked Questions About NAEP Sampling." *EdSource*, 2005.

Parker, Verilette, and Brian Gerber. "Performance-Based Assessment, Science Festival Exhibit Presentations, and Elementary Science Achievement." *Journal of Elementary Science Education* 14, no. 1 (Spring 2002): 59-67.

Phelps, Richard P. "Teach to the Test?" *The Wilson Quarterly (1976-)* 35, no. 4 (Fall 2011). http://www.jstor.org/stable/41484371.

POLICY RECOMMENDATIONS FOR NEW YORK CITY'S NEXT MAYOR. New York City: New York City Bar, n.d.

———. "Racial and Ethnic Achievement Gap Trends: Reversing the Progress toward Equity?" *Educational Researcher* 31, no. 1 (January/February 2002): 3-12.

Raising standards or raising barriers? Inequality and high-stakes testing in public education (Gary Orfield, Mindy L. Kornhaber)

Ravitch, Diane, and Tom Loveless. "Broken Promises: What the Federal Government Can Do To Improve American Education." *Brookings Institution*, March 1, 2000.

Resnick, Lauren. "2009 Wallace Foundation Distinguished Lecture: Nested Learning Systems for the Thinking Curriculum." *Educational Researcher* 39, no. 3 (April 2010): 183-97.

Rhodes, Jesse H. *An Education in Politics: The Origins and Evolution of No Child Left Behind*. N.p.: Cornell University Press, 2012.

Robenstine, Clark. "Confronting the Assumptions That Dominate Education Reform." *Journal of Thought* 32, no. 3 (Fall 1997): 113-21.

Robinson, Ken. *Creative Schools: The Grassroots Revolution That's Transforming Education*. N.p.: Penguin Books, 2015.

Sacks, Peter. "Standardized Testing: Meritocracy's Crooked Yardstick". Change, vol. 29, o. 2, 1997, pp. 24–31.

Schellenberg, Rita, and Timothy Grothaus. "Promoting Cultural Responsiveness and Closing the Achievement Gap with Standards Blending." *Professional School Counseling* 12, no. 6 (August 2009): 440-49.

Schumer, Robert. "Finland not an apt educational model for U.S. schools." *The Star Tribune*, July 11, 2014. http://www.startribune.com/finland-not-an-apt-educational-model-for-u-s-schools/266823501/.

Silva, Elena. "Measuring Skills for 21st-Century Learning." *The Phi Delta Kappan*, May 2009, 630-34.

Stanford School Redesign Network. *What is Performance-Based Assessment?* N.p., 2008.

Stanford, Jason. "Mute the Messenger When Dr. Walter Stroup showed that Texas' standardized testing regime is ?awed, the testing company struck back." *Texas Observer*, September 3, 2014. https://www.texasobserver.org/walter-stroup-standardized-testing-pearson/.

Starr, Joshua P., and Margaret Spellings. "Examining High-Stakes Testing." In *Education Next*. Previously published in *Education Next*, Winter 2014, 70-77.

Stedman, Lawrence. "The Sandia Report and U.S. Achievement: An Assessment." *The Journal of Educational Research* 87, no. 3 (January/February 1994): 133-46.

Summary of the Every Student Succeeds Act, Legislation Reauthorizing the Elementary and Secondary Education Act. N.p.: National Conference of State Legislatures, n.d.

TASB Legal Services, comp. *Update on the Every Student Succeeds Act (ESSA): Navigating through Public School Accountability after ESSA*. Austin, TX: Texas Association of School Boards, n.d.

Timar, Thomas B. "The Institutional Role of State Education Departments: A Historical Perspective." *American Journal of Education* 105, no. 3 (May 1997): 231-60.

Tourangeau, Karen, Christine Nord, Than Le, Kathleen Wallner-Allen, Nancy Vaden-Kiernan, Lisa Blaker, Michelle Najarian, and Gail Mulligan. *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11*. N.p.: National Center for Education Statistics, 2011.

U.S. Department of Education American Recovery and Reinvestment Act Report: Summary of Programs and State-by-State Data. N.p.: U.S. Department of Education, 2009.

U.S. Department of Education. "EVERY STUDENT SUCCEEDS ACT Assessments under Title I, Part a & Title I, Part B: Summary of Final Regulations." In *U.S. Department of Education*.

Uzzell, Renata, Nkemka Anyiwo, Moses Palacios, Candace Simon, Ray Hart, and Michael Casserly, comps. *BEATING THE ODDS Analysis of Student Performance on State Assessments*. N.p.: Council of the Great City Schools, 2012.

Author Biography

Christina Breitbeil was born in Houston, Texas on November 9, 1994. She enrolled in the Plan II Honors program at the University of Texas at Austin in 2013, with double majors in Plan II and English. In college, she wrote for the University's campus newspaper, *The Daily Texan*, edited and served on the executive board for the *Texas Undergraduate Research Journal*. She was also a member of the University's oldest honorary service and leadership organization, the Texas Orange Jackets. Outside of school, she worked as an Editorial Intern and wrote for *RealClearPolitics* in Washington D.C., and worked as a Legislative Intern in the Texas House of Representatives. She graduated Phi Kappa Phi, recipient of both the J.J. Jake Pickle Citizenship Award and the Randy Diehl Prize in Liberal Arts, in 2017. Ms. Breitbeil will begin teaching elementary school in Brooklyn, New York this fall, while she earns her Masters in Teaching.